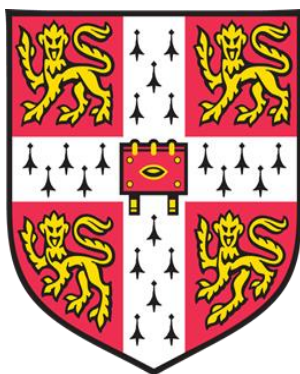


Computer-Aided Design of Ligands at Multiple Protein
Targets
for Multifactorial Diseases

Leen Kalash

Murray Edwards College
University of Cambridge



The present thesis is hereby submitted for the degree of
Doctor of Philosophy

December 2018

Disclaimer

The present thesis is the result of my own work except where explicitly stated in the text. The document length does not exceed the word limit set by the Degree Committee.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Dr. Andreas Bender and Professor Robert Glen for their help, guidance, and supervision in my projects throughout my PhD.

Furthermore, I would like to thank Dr. Graham Ladds, Professor Rudolf Allemann, and Dr. Eddy Sotelo for the scientific collaborations, and Dr. Natalia Aniceto for proof reading parts of the thesis.

Last but not the least, my sincere thanks to my parents Sana and Nabeeh, my sisters Maya and Raya, and my brother-in-law Hani, for their continuous support, motivation and inspiration throughout my PhD.

List of Publications

L Kalash, J Cresser-Brown, J Habchi, C Morgan, D J. Miller, R Glen, R Allemann, A Bender
Structure-based design of allosteric calpain-1 inhibitors populating a novel bioactivity space. -
European Journal of Medicinal Chemistry. (2018) 157, 1264
DOI: 10.1016/j.ejmech.2018.08.049

This publication is described in Chapter 2.

L Kalash, C Val, J Azuaje, MI Loza, F Svensson, A Zoufir, L Mervin, G Ladds, J Brea, R Glen,
E Sotelo, A Bender Computer-aided design of multi-target ligands at A₁R, A_{2A}R and PDE10A,
key proteins in neurodegenerative diseases. – Journal of Cheminformatics (2017) 9, 67

DOI: 10.1186/s13321-017-0249-4

This publication is described in Chapter 3.

L Kalash, I Winfield, D Safitri, M Bermudez, S Carvalho, R Glen, G Ladds, A Bender
Structure-based identification of dual ligands at the A_{2A}R and PDE10A with anti-proliferative
effects upon lung carcinoma cell-lines (In Prep)

This manuscript is described in Chapter 4.

LH Mervin, KC Bulusu, L Kalash, AM Afzal, F Svensson, MA Firth, I Barrett, O Engkvist, A
Bender Orthologue chemical space and its influence on target prediction. – Bioinformatics
(Oxford, England) (2017) 34, 72

DOI: 10.1093/bioinformatics/btx525

Parts of this publication are described in Chapter 1.

N Mateu, SL Kidd, L Kalash, HF Sore, A Madin, A Bender, DR Spring Synthesis of
structurally diverse N-substituted quaternary carbon containing small molecules from α,α -
disubstituted propargyl amino esters. – Chemistry A European Journal (2018) 24, 13681

DOI: 10.1002/chem.201803143

CP Baburajeev, CD Mohan, V Pandey, S Rangappa, N Shivalingegowda, L Kalash, S
Devaraja, A Bender, PE Lobie, KS Rangappa, Basappa Synthesis of C-C, C-N Coupled Novel
Substituted Dibutyl Benzothiazepinone Derivatives and Evaluation of Their Thrombin
Inhibitory Activity. – Bioorganic Chemistry (2019) 87, 142

DOI: 10.1016/j.bioorg.2019.03.004

Summary

Today, drug discovery predominately focuses on the design of ligands with high selectivity towards a specific biological target. A significant limitation in the case of multi-factorial diseases (e.g. neurodegenerative disorders) is that effective therapy may require multi-target drugs addressing the complexity of multi-factorial pathologies. Here, single- and multi-target ligand design was investigated to discover novel compounds active at multiple proteins/multiple binding sites including allosteric ligands.

Calpain-1, a challenging target, was selected to develop and evaluate computational approaches to the discovery of novel ligands. Current selective calpain-1 inhibitors are reported to bind to an allosteric site and their mode of action has remained elusive. To elucidate this, a structure-based virtual screening protocol was implemented to find chemically novel compounds with improved selectivity and a reduced side-effects profile.

To develop methods for the discovery of multi-target ligands, a multi-target design approach, which could be beneficial in the treatment of Lung Carcinoma and Neurodegenerative diseases, was investigated. A novel ensemble of proteins was targeted to elevate intracellular cAMP, deemed to be beneficial in these diseases resulting in the discovery of ligands with high binding affinity at three targets, PDE10A, A₁ and A_{2A} receptors.

In tandem, functional activity at the A_{2A} receptor and PDE10A was investigated, resulting in the discovery of novel compounds, which exhibited anti-proliferative effects in lung carcinoma cell-lines correlating with the co-expression of the two targets and increased cAMP levels. Critically, the dynamics of one amino acid residue, Val₈₄, was identified as a novel conformational descriptor of A_{2A} receptor activation.

Overall, novel single- and multi-target ligand design approaches are presented in this work, which could be applicable to a wide range of ligand design problems, across (multi-factorial) disease areas and target families. The findings may facilitate improved design of allosteric calpain-1 inhibitors using the PEF(S) domain, and encourage investigating the therapeutic benefits of dual ligands at the A_{2A} receptor and PDE10A against lung cancer *in vivo*.

Table of Contents

Disclaimer.....	ii
Acknowledgements	iii
List of Publications	iv
Summary.....	v
Table of Contents	vi
1 Introduction	1
1.1 Drug discovery.....	1
1.1.1 Tuning ligand selectivity in drug discovery	1
1.1.2 The transition from single-target to multi-target drug approach for the treatment of multi-factorial diseases	3
1.1.3 Current status of multi-target drugs in relation to other therapeutic strategies ...	3
1.1.4 Examples of clinical success of multi-target drugs encourages movement towards multi-target ligand design.....	5
1.2 Rational design of multi-target drugs.....	7
1.2.1 Existing challenges in selecting multi-target combinations in drug discovery ...	7
1.2.2 Reported Strategies for addressing the challenge in selecting the ‘right’ target combination for multi-target ligand design	8
1.2.3 Promising target combinations for multi-target drug design (focusing on GPCRs and Enzymes)	9
1.2.4 A novel multi-target combination- modulation of A ₁ and A _{2A} receptors and inhibition of the enzyme PDE10A.....	11
1.2.5 Current computational approaches for the design of multi-target drugs	13
1.3 Aims of this work.....	35
2 Structure-based design of allosteric calpain-1 inhibitors populating a novel bioactivity space.....	37
2.1 Introduction	37
2.2 Materials and Methods.....	42
2.2.1 Extraction and preparation of purchasable compounds for structure-based virtual screening against PEF(S).....	42
2.2.2 Receptor Preparation of PEF(S)	42
2.2.3 Cut-off generation for compound selection from docking model	42

2.2.4	Docking	44
2.2.5	Multi-Dimensional Scaling (MDS) analysis of the shortlisted compounds 1-10.....	44
2.2.6	Experimental validation of the virtual screening protocol	45
2.3	Results and Discussion.....	50
2.3.1	Structure-based virtual screening of purchasable ligands against PEF(S)	50
2.3.2	MDS plot shows that shortlisted PEF(S) binders occupy a novel region in chemical space	50
2.3.3	FRET based Inhibition Assay.....	52
2.3.4	TNS Displacement Assay.....	56
2.3.5	Analysis of molecular docking studies of representative calpain-1 inhibitors 1 and 10, and compounds 2-5.....	57
2.3.6	Computational assessment of CNS permeability for representative calpain-1 inhibitors 1 and 10	60
2.4	Conclusions	61
3	Structure- and ligand-based design of multi-target ligands at A ₁ R, A _{2A} R and PDE10A-key proteins in neurodegenerative diseases.....	62
3.1	Introduction	62
3.2	Materials and Methods.....	65
3.2.1	Method for the selection of reference molecules for the design of multi-target ligands... ..	65
3.2.2	Designing new multi-target ligands	66
3.2.3	Target prediction.....	66
3.2.4	Receptor preparation	68
3.2.5	Ligand Preparation.....	68
3.2.6	Cut-off generation for compound selection from docking models.....	68
3.2.7	Docking	71
3.2.8	Substructural analysis	71
3.2.9	Experimental validation of 2-aminopyridine-3-carbonitriles as multi-target ligands at A ₁ R, A _{2A} R, and PDE10A	72
3.3	Results and Discussion.....	80
3.3.1	Design of synthetically feasible A ₁ R/A _{2A} R-PDE10A multi-target ligands	80
3.3.2	Target prediction of the designed RECAP library	80

3.3.3	Docking of the compounds predicted as A ₁ R/A _{2A} R-PDE10A multi-target ligands...	81
3.3.4	Substructure analysis of the compounds predicted as A ₁ R/A _{2A} R-PDE10A multi-target ligands.....	81
3.3.5	Synthesis of novel 2-aminopyridine-3-carbonitriles.....	82
3.3.6	Pharmacological evaluation of novel 2-aminopyridine-3-carbonitriles.....	82
3.3.7	(SAR) Structure-activity relationship analysis	87
3.3.8	Compound selectivity assessment	92
3.3.9	Analysis of the predicted binding modes of the synthesized 2-aminopyridine-3-carbonitriles	97
3.3.10	Computational assessment of CNS permeability	100
3.4	Conclusions	100
4	Structure-based identification of dual ligands at A _{2A} R and PDE10A with anti-proliferative effects upon lung carcinoma cell-lines.....	102
4.1	Introduction	102
4.2	Materials and Methods.....	105
4.2.1	Design approach of the dual ligands at the A _{2A} R and PDE10A.....	105
4.2.2	Selection of the A _{2A} R protein crystal structure for shortlisting triazoloquinazoline candidates as A _{2A} R agonists	105
4.2.3	Ligand Preparation.....	107
4.2.4	Receptor Preparation.....	107
4.2.5	Enrichment of agonists by the A _{2A} R docking model (PDB ID: 2YDO)	107
4.2.6	Cut-off generation for compound selection as candidates of A _{2A} R agonists from the docking model.....	108
4.2.7	Docking	108
4.2.8	MD Simulations.....	109
4.2.9	Experimental validation of the virtual screening protocol	110
4.3	Results and Discussion.....	113
4.3.1	Method of selecting triazoloquinazolines as candidates of dual ligands at A _{2A} R and PDE10A	113
4.3.2	Analysis of the predicted binding modes of representative triazoloquinazolines 36-39 shortlisted for experimental validation.....	114

4.3.3	MD simulation suggests that the conformational change of the His ₂₅₀ residue contributes to shaping the orthosteric site pocket to favor selectivity for A _{2A} R agonists.....	116
4.3.4	MD Simulation analysis reveals the shift in Val ₈₄ as a requirement for receptor activation by A _{2A} R ligands.....	119
4.3.5	Confirmation of triazoloquinazolines as PDE10A inhibitors.....	122
4.3.6	Experimental validation of triazoloquinazolines as A _{2A} R agonists.....	124
4.3.7	Mammalian validation	127
4.3.8	Dual PDE10A inhibition and A _{2A} R agonism is anti-proliferative in CHO-K1-A _{2A} R cells.....	129
4.3.9	Dual PDE10A inhibition and A _{2A} R agonism is anti-proliferative in Lung carcinoma cell-lines	132
4.4	Conclusions	141
5	Conclusions & Future Work.....	142
	References	144
	Appendix A.....	172
	Appendix B.....	173
	Appendix C.....	174
	Appendix D.....	175
	Appendix E.....	176

1 Introduction

1.1 Drug discovery

1.1.1 Tuning ligand selectivity in drug discovery

A contemporary approach in drug discovery has been the design of potent ligands with high selectivity towards a target associated with a particular disease of interest, followed by structural optimization for drug development.[1] Generally, undesirable side effects might arise when ligands interact with e.g. other targets, DNA, RNA, lipids, sugars, metabolites.[2] Therefore, huge effort has been made to measure the off-target interactions that might lead to adverse side effects such as those with ion channels[3], transporters and cytochrome P450s (CYPs) etc.[4, 5] In addition, research has been directed to achieve family or subtype selectivity for homologous targets binding to similar native substrates. For instance, this has been a challenge in the kinase family (phosphotransferases), where each kinase could bind to its co-factor ATP while transferring a phosphate group to a substrate.[6] Hence, one of the aims was to find a ligand that could either inhibit selectively one or alternatively a group of kinases that share a particular biochemical pathway of interest, in order to avoid the undesirable side effects that might arise.[7] Thus, it has generally been achievable to tune the selectivity of ligands to a subset of kinases rather than a single kinase.[8]

In practice, it is challenging to design a drug exhibiting narrow selectivity over undesirable targets. It would be essential then, to develop an understanding of the factors driving selectivity towards any particular target of interest, and in turn this would allow for more effective design of ligands with the desired selectivity.[9–11] This gives rise to the need for rational approaches to tune ligand selectivity. One of the currently existing computational approaches for designing selective ligands is proteochemometrics (PCM) used for simultaneously modeling bioactivity of a group of ligands against a set of related protein targets.[12] In addition, structure-based approaches are generally employed to enhance ligand selectivity by improving the shape and electrostatic complementarities between the ligand and the binding pocket of the protein, as well as addressing issues such as conformational selection and the flexibility of proteins.[2]

Designing *allosteric* ligands which target a site specific to the protein of interest but does not include all of, or is distinct from, the site occupied by e.g. a substrate or ligand, usually termed the active or the orthosteric site, would be one way to approach this problem.[13] Allosteric binding, through transmission of conformational effects, can in some cases inhibit or activate a protein mediated mechanism.[14] For instance, Hemoglobin (Hb) is one classical example of

allostery where a protein, which is a subunit of a multi-subunit protein, contributes to regulating the physiological properties of Hb allosterically.[14]

To illustrate this approach, one of the challenges of designing selective ligands has been addressed in this work by selecting calpain-1, one of the dimeric calpains, as a protein target. Calpain-1 constitutes a promising therapeutic target for many diseases e.g. in cardiovascular, neurodegenerative and ischaemic diseases.[15, 16] Most compounds that target the *active* site of calpain-1 inhibit a broad spectrum of cysteine proteases, thereby resulting in undesirable side effects.[17] Selective calpain-1 inhibitors, such as PD150606, which included a specific α -mercaptoacrylic acid sub-structure, were reported to bind to an allosteric site of calpain, the penta-EF hand calcium binding domain, PEF(S), but their mode of action has remained unclear. [18]. Here, the crystal structure of PEF(S) was used in virtual screening to discover allosteric PEF(S) binders that populate a novel chemical space to assist in elucidating their mode of action while having promising selectivity and a reduced side-effects profile (Figure 1).

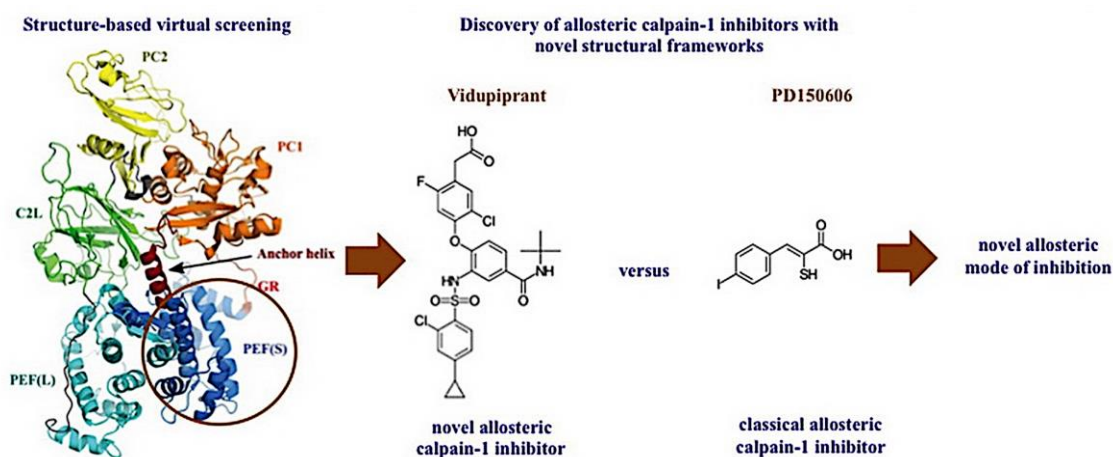


Figure 1. The use of the crystal structure of the penta-EF hand calcium binding domain, PEF(S) of calpain-1 in virtual screening to discover PEF(S) binders that populate a novel chemical space, and elucidation of a novel allosteric mode of action with promising selectivity and a reduced side-effects profile.

1.1.2 The transition from single-target to multi-target drug approach for the treatment of multi-factorial diseases

The design of selective ligands[1, 2] has been proven successful for diseases with well-defined mechanism and pathophysiology. In the case of these diseases, multi-target drugs were undesirable and have been long associated with adverse side effects.[1, 2] In contrast, for multi-factorial diseases, such as neurodegenerative disorders, inflammatory diseases, and cancer that exhibit more-complex pathological mechanisms due to an ensemble of factors (e.g. genetic and environmental) contributing to the severity of the diseases,[19–28] a single target approach often fails to demonstrate reliable therapeutic effects.[29–32] This gives rise to the need for multi-target drugs, which could address the complexity of multi-factorial diseases with improved therapeutic efficacy. Additionally, they may exhibit additive or synergistic effects. This could result then in better safety profiles in comparison to single-target drugs due to the lower dosage requirements and improved efficacy.[33]

On this basis, multi-target drug design has emerged as a new paradigm in drug discovery, with more research being directed towards finding novel and effective multi-target drugs particularly for the treatment of multi-factorial diseases. In confirmation of this approach, many marketed multi-target drugs are already available exemplifying the success of this methodology. These drugs exhibit complex pharmacological profiles.[34] For instance, in the case of cancer and infectious diseases, the choice of drugs active at multiple targets may be attributed to the emergence of resistance in e.g. cancer or bacterial infection, resulting from genetic mutations or in the case of anxiolytic drugs to the requirement to modulate multiple receptors in the central nervous system (CNS).[35] This suggests that specifically designing multi-target drugs is a promising approach to treat multi-factorial diseases.

1.1.3 Current status of multi-target drugs in relation to other therapeutic strategies

An analysis was conducted by Ramsay *et al*[35] for the US Food and Drug Administration (FDA) of approved new molecular entities (NMEs) over the period of 2015 to 2017. This analysis aimed to showcase the success of multi-target drugs in the clinic at that time, and was similar to a previous analysis done by Lin *et al*. [36] Information about the currently approved drugs was extracted from the DrugBank database, which includes drug target(s) and mechanism(s) of action (MoA), and they were subdivided into single-target and multi-target drugs. The major highlight of the analysis is shown in Figure 2, illustrating that the number of single-target drugs (34%) is greater than that of multi-target drugs (21%); however, the latter

have constantly increased over a period of years (16%).[37] In addition, combining the 21% of multi-target drugs and newly approved therapeutic combinations (10%), the total percentage of the two approaches is 31%, which is approaching that of single-target drugs (34%). This clearly reveals the rising interest in polypharmacological approaches in certain therapeutic areas, in particular multifactorial diseases such as CNS disorders, cancer, and infectious diseases.[35]

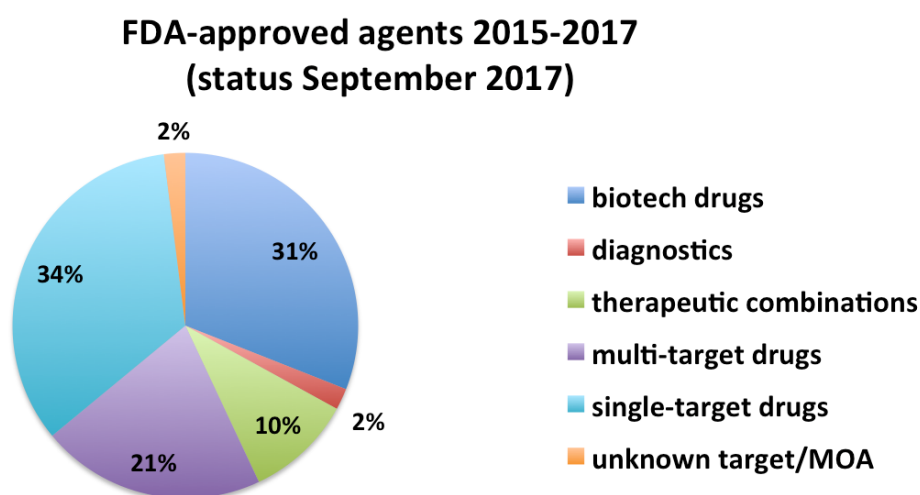


Figure 2. Analysis was done by Ramsay *et al*[35] for FDA approved new molecular entities (NMEs) from 2015 to 2017. Information about currently approved drugs was extracted from the DrugBank database, which includes drug target(s) and mechanism(s) of action (MoA). They were subdivided into single-target and multi-target drugs. The major highlight of the analysis was that despite the fact that the number of single-target small molecules (34%) is more than that of multi-target drugs (21%), the latter has consistently increased (16%).

An alternative polypharmacological approach to multi-target drugs is combination therapy, which uses drugs with different modes of action for the treatment of complex diseases.[34] The percentage distribution of multi-target drugs and therapeutic combinations approved in 2015–2017, according to the Anatomical Therapeutic Chemical (ATC) Classification System, is depicted in Figure 3. Generally, multi-target drugs are much more widely used as compared to combination therapies across all disease areas, in particular as anti-neoplastic agents and for the treatment of nervous system disorders.[35]

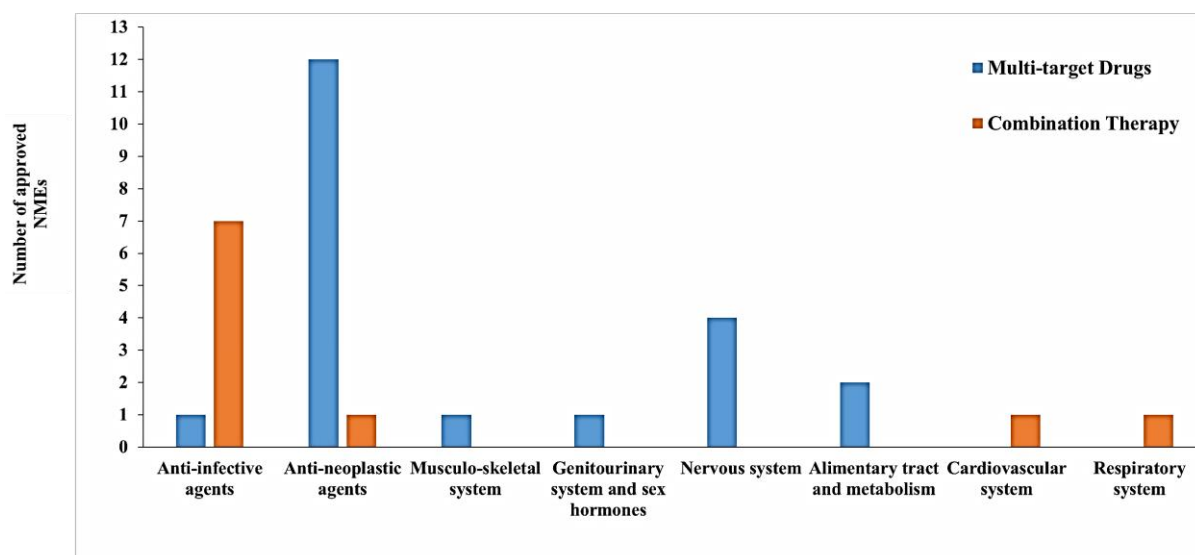


Figure 3. Distribution of multi-target drugs and therapeutic combinations approved in 2015–2017 according to the Anatomical Therapeutic Chemical (ATC) Classification System reveals that generally, multi-target drugs are much more widely used as compared to combination therapies across all disease areas, in particular as anti-neoplastic agents and for the treatment of nervous system disorder[35].

In fact, the use of multi-target drugs exhibiting multiple biological properties may have advantages over combination therapies, where different challenges are encountered. The first is coping with several drugs with different bioavailabilities, pharmacokinetic properties (PK), and metabolism, which may not be the case when administering a single multi-target drug with multiple modes of action on several targets into the body.[34] In addition, pre-clinical development of a multi-target drug, which includes PK and ADMET (absorption, distribution, metabolism, and excretion – toxicity) optimization, is similar to that of single-target drugs. Thus the multi-target drug approach may be less complicated than combination therapies. One major advantage is that undesirable drug–drug interactions would not be encountered in the case of multi-target drugs.[34] Hence, from these considerations, the multi-target drug approach may be a more efficient and cost-effective poly-pharmacological option as compared to drug combinations. On the other hand, many challenges are encountered with multi-target drugs, which are mentioned in section 1.2.1.

1.1.4 Examples of clinical success of multi-target drugs encourages movement towards multi-target ligand design

The multi-target approach appears to be promising, in particular for the treatment of neurodegenerative disorders, given their complexity and multifactorial nature.[38] For example, rasagiline is a drug that was approved in 2006 for the treatment of neurodegenerative

diseases, and it exhibits a multi-target profile. In addition, safinamide is the first multi-target drug approved for neurodegeneration for more than a decade.[39] It was originally developed as an anticonvulsant agent and recently in March 2017 it was approved as an adjunctive treatment for Parkinson's disease due to its multi-target nature. It is thought that this drug confers neuroprotective effects by combining monoamine oxidase-B inhibition with activity at sodium-gated ion channels with the release of glutamate, which in turn controls motor symptoms.[40] In addition, recently an increased number of multi-target drugs have been developed to treat schizophrenia and major depressive disorders.[38]

For example, antagonizing three targets simultaneously, namely D₂, serotonin 5-HT_{2A} and α 1-adrenergic receptors has been a common approach used for the treatment of schizophrenia.[41] This generally aimed to improve antipsychotic efficacy and minimize adverse effects. For this purpose, the arylpiperazine substructure was used for analogue modification in order to tune a fine balance of receptor activities at D₂, 5-HT_{1A} and 5-HT_{2A}. As a result, aripiprazole, was released into the market in 2015, as the first modulator of serotonin/dopamine activity. However, it exhibited undesirable side effects, reported to be related to its interaction with post-synaptic D₂ receptors. Subsequently, brexpiprazole was approved on October 2015, as a novel D₂ and 5-HT_{1A} partial agonist. In comparison to aripiprazole, it exhibited reduced inherent activity at D₂ receptors and more balanced activities at 5-HT_{2A}, 5-HT_{1A}, and the α 1B receptor subtypes.[41] Cariprazine is also another FDA approved drug that modulates serotonin/dopamine activity, also having an arylpiperazine substructure. It was marketed in September 2015 for the treatment of schizophrenia and bipolar disorders. Being a partial agonist of the D₂ and D₃ receptors, this mode of action is novel compared to other antipsychotics, which are D₂ and 5-HT_{2A} partial agonists.[42]

Besides neurodegenerative disorders, another key multi-factorial disease where polypharmacological approaches are the most relevant is cancer. Several proteins and pathways contribute to the complexity of this disease, often leading to the progression of tumor growth.[43–45] Inhibiting multiple kinases and pan-inhibitors of histone deacetylases (HDACs) have recently been the most investigated polypharmacological approaches for cancer treatment. For example, lenvatinib is a reversible inhibitor of multi-tyrosine kinase receptors that simultaneously modulates their activities (vascular endothelial growth factor receptors (VEGFR), fibroblast growth factor receptors (FGFR), RET, mast/stem cell growth factor receptor kit (SCFR), and platelet-derived growth factor receptor (PDGFR) beta). These receptors are associated with pathogenic angiogenesis, tumor growth, and cancer

progression.[46] In addition, due to its multi-target activity profile, lenvatinib was approved for the treatment of radioiodine-refractory thyroid cancers. Neratinib is another multi-tyrosine kinase inhibitor, which is irreversible, and it targets epidermal growth factor receptor (EGFR), and human epidermal growth factor receptor 2 (HER2), receptors that are highly expressed in several carcinomas, and are thought responsible for its anti-tumor activity.[47]

Another common approach to multi-kinase inhibition includes dual inhibitors of cyclin-dependent kinase (CDK) 4 and 6, which have been approved by the FDA for breast cancer therapy e.g. palbociclib, abemaciclib and ribociclib.[48] In addition, midostaurin, is a multi-kinase inhibitor derived from the pan kinase inhibitor staurosporine. It was approved on April 2017 for the treatment of acute myeloid leukemia *via* the inhibition of the activities of protein kinase C alpha (PKC alpha), VEGFR2, KIT, PDGFR and WT and/or mutant FLT3 tyrosine kinases.[49]

In summary, there are several examples of clinically successful multi-target drugs and this suggests that more effort towards the development of multi-target ligand design would be beneficial. Examples such as these have inspired the implementation of multi-target ligand design approaches documented in this work, with the objective of further exploring the potential therapeutic benefits in multi-factorial diseases.

1.2 Rational design of multi-target drugs

1.2.1 Existing challenges in selecting multi-target combinations in drug discovery

Based on an analysis of FDA-approved new molecular entities (NMEs) from 2015 to 2017,[35] highlighting the increasing number of marketed multi-target drugs, it appears that extending the applicability of multi-target drugs across different therapeutic areas would allow the discovery of more effective treatments. It remains a challenging task however, to rationally design multi-target drugs.[50–52]

The first challenge lies in selecting the right combination of targets for a particular disease. Despite the availability of many online resources[53–55] which can be used to identify possible targets, target selection is often ambiguous. For this reason, it is essential to develop an in-depth understanding of the relationships between target(s) and disease(s), the associations between key metabolic pathways, which targets, are key, which drugs have the correct therapeutic profile, and which diseases are druggable by this approach in the absence of potential adverse events.[56] Moreover, it is highly beneficial to achieve additive or synergistic

effects.[57]. Hence, to achieve all these simultaneously requires the design of multi-target compounds with a particular functional activity profile.[58]

In most cases multi-target drug design has historically been based on a fairly simple approach, that of merging two distinct pharmacophores derived from sets of active compounds, where each compound is initially selective towards its corresponding target, into the design of a single compound[59–61]. Based on the degree of similarity of the starting structural frameworks of each compound, pharmacophores could either be fused or merged. In contrast, if they are sufficiently different, then pharmacophores could be conjugated with cleavable or non-cleavable linkers. This strategy however could result in compounds with poor properties as drugs e.g. poor pharmacokinetics, which may not be suitable for acute or chronic diseases.[62] The nature of the targets involved and the availability of starting frameworks are both important parameters to be considered when generating multi-target compounds. It is also important that each framework retain interactions with specific target (i.e. those compounds from which the fragment was derived).[52] This is quite challenging to achieve and requires structure–activity relationships of the starting compounds with their corresponding targets, especially if these targets are either slightly related or completely unrelated e.g. belonging to different protein classes.[52]

Other critical factors that need to be addressed include modulating each target to the required degree and avoiding any off-target effects. In particular, the latter aspect is of high importance in designing multi-target compounds against same protein family. Examples are multi-kinase inhibitors, or those of shared functional domains and/or binding sites across target families.[52] It is evident then that the rational design of multi-target compounds is challenging - the right choice of target combination is key while achieving balanced activity towards each target without any off-target activity whilst keeping drug-like properties.

1.2.2 Reported Strategies for addressing the challenge in selecting the ‘right’ target combination for multi-target ligand design

From the literature, network graphs are one approach that can help to obtain insight into the association of protein targets with multi-factorial diseases, and reveal a mechanism that can explain how perturbations in cellular networks might result in certain phenotypes, including disease. Hence, this is one approach to guide the selection of the ‘right’ combination of targets to find therapies for complex diseases. The approach is to model metabolic and signaling pathways by mining high-throughput experimental datasets and integrating the data into

network models. This can assist in modeling outcomes of drug intervention and also in understanding the complex interactions between multi-target drugs and their cellular targets with the objective of explaining their efficacy and potential side effects.[63–70] A better understanding of the mode of action of multi-target drugs can then be achieved by focusing on the more promising targets in the disease networks where inhibition can influence the development of disease phenotypes such as cancer progression.[63, 70–73] In addition, this may allow better understanding of drug resistance and side effects, for example by highlighting particular targets or how their pathways are interconnected.[74–79] The robustness of disease networks can be used to explain the efficacy of multi-target drugs, for example by perturbing a specific subset of nodes in the network that are key to phenotypic development. Network-based drug discovery is therefore one approach to search for target combinations that represent nodes in disease networks that could potentially be perturbed by multi-target drugs, enabling the development of more effective and safer therapies.

1.2.3 Promising target combinations for multi-target drug design (focusing on GPCRs and Enzymes)

G protein-coupled receptors (GPCRs) have been widely studied as drug targets due to their pharmacological tractability and strong association with human pathophysiology. Around 34% of all FDA approved drugs are GPCR targeted.[80, 81] Since GPCRs are in general druggable receptors, GPCR targeted ligands are appealing as part of multi-target compounds. For instance, histamine receptor antagonism was combined to cholinesterase and monoamine oxidase inhibition in a drug for Alzheimer's disease.[82]

However, defining the precise mode of action and mechanism can be challenging. The mode of action of GPCRs can be modulated by multiple signaling pathways[83], including functional selectivity (biased agonism).

Over the past twenty years, a significant number of receptors have been characterized by X-ray crystallography, including membrane-bound receptors such as GPCRs. This has been made possible by significant progress in cloning and purification of membrane-bound proteins and advances in structure solution and refinement. In fact, the currently available data on the structure and function of GPCRs evolved from homology modeling of related GPCRs such as rhodopsin (class A GPCRs) and has since been expanded to include X-ray structures of GPCR receptor subtypes belonging to other classes (B, C, D, E and F that don't share any sequence homology among each other).[84–86] It is challenging to design GPCR ligands and understand

the mechanism by which they elicit agonism or antagonism. This requires an analysis of the complex dynamics of the system including allosteric effects. For this reason, MD simulation and analysis of the ligand-protein complexes is a promising approach.[87] A challenging aspect that needs to be considered in the design of GPCR ligands is the relatively large degree of flexibility exhibited by GPCR proteins. To take this into account, this can be partially addressed by docking methods that consider receptor flexibility (discussed in section 1.2.5.3 Docking methodologies). To develop a better understanding of signaling in cells and *in vivo*, efforts have been made to explore transient dimerization[88], heterodimers,[89] and internalized megamers[90] with a view to assist in the design of selective ligands that would lead to various aggregation states.

Enzymes play a key role in life processes and in pathophysiology, and they are also key drug targets. Approximately half of existing drugs are active against enzymes and recent studies of the human genome suggest that enzymes constitute the majority of druggable targets. Enzymes are therefore key targets for a number of diseases and the design of modulators of enzyme activity e.g. inhibitors continues to be an important avenue of research in drug discovery and development.[91]

In fact, the active site of an enzyme and if present, allosteric pockets, may be suited for interactions with drug-like inhibitors. In comparison to GPCRs, it is often more straightforward to design inhibitors that compete with the native substrate (competitive inhibitors). Hence, targeting an enzyme active site is generally seen as an attractive approach in drug design. This is assisted by availability of many enzyme crystal structures. However, focusing on reversible/competitive active site inhibitors may not be the best approach in all cases. For example, effective drugs could be non-competitive or could be irreversible inhibitors of enzyme activity[91, 92].

An important measure of enzyme activity in the presence of an inhibitor is to study its kinetics. Two frequently used tools for the evaluation of binding to purified enzymes are surface plasmon resonance and isothermal titration calorimetry (ITC). In contrast, in GPCRs, it is important to measure functional effects, rather than binding.[93]

Many successful examples of multi-target ligands targeting both enzymes and GPCRs have been reported in the literature. Examples include dual ligands of Cannabinoid CB₁R inverse agonists and acetylcholine (AChE) inhibitors, μ -opioid receptor (MOP) agonists and nitric oxide synthase (NOS) inhibitors for the treatment of pain, and norepinephrine (NER)

antagonists and NOS inhibitors for the treatment of neuropathic pain.[94] Another example is that of compounds acting as A_{2A} adenosine receptor antagonists and monoamine oxidase B (MAO B) inhibitors;[95] a promising target combination for the treatment of Parkinson's disease. Additional examples include dual-active ligands of histamine H₃R antagonists and AChE inhibitors, which have been suggested to be beneficial in treating cognitive disorders.[94] Given all of these examples, it would be beneficial to design multi-target ligands that simultaneously target specific enzymes and receptors as this may produce improved therapeutic benefit in multi-factorial diseases.[94]

1.2.4 A novel multi-target combination- modulation of A₁ and A_{2A} receptors and inhibition of the enzyme PDE10A

Cyclic-AMP (cAMP) is involved in many biological processes such as cell growth and adhesion, neuronal signaling, immune function, and metabolism etc. As a second messenger involved in multiple signaling pathways, in specific instances and in particular disease states, elevation of intra-cellular cAMP concentrations has demonstrated therapeutic benefit in multi-factorial diseases such as CNS traumas, autoimmune disease, inflammatory diseases, and cancer.[96–98] For example, intracellular cAMP levels have been shown to have both pro-[99–101] and anti-proliferative effects,[101–106] depending on the cancer cell type. cAMP levels are spatially and temporally coordinated (e.g. by specifically positioned PDE4 enzymes), creating concentration gradients within the cell to elicit specific actions and outcomes[107].

The intracellular elevation of cAMP can be achieved by various ligands that preferentially target single receptors, such as Bay K8644 (a calcium channel agonist), TTX (a sodium channel antagonist) and H89 (a protein kinase A inhibitor). Other examples of compounds that elevate cAMP include KT5720 (a protein kinase A (PKA) inhibitor), LY29400 (an inhibitor of phosphoinositide 3-kinases (PI3Ks)), PD98059 (a MAPK kinase inhibitor), tyrphostin AG490 (a tyrosine kinase inhibitor), myelin-associated inhibitors, and Forskolin.[33, 108–112] Inhibition of phosphodiesterases also results in maintenance of elevated intracellular cAMP concentrations; examples of known inhibitors include IBMX, a non-competitive selective phosphodiesterase inhibitor, and rolipram, a selective phosphodiesterase-4 inhibitor.[113, 114] Additionally, the modulation of adenosine A₁ and A_{2A} receptors contribute to the elevation of cAMP levels.[115, 116]

One particular enzyme of interest in this work is the phosphodiesterase 10A (PDE10A), which plays a role in neurodegenerative,[117–119] inflammatory[120] and cancer-related diseases.[98] Inhibition of PDE10A resulting in maintenance of elevated intracellular cAMP concentrations was suggested to be effective in the treatment of these diseases (Figure 4A. adapted and modified from Lee *et al*). Similarly, the modulation of two GPCRs the adenosine A_1 and A_{2A} receptors (A_1R and A_{2AR}), plays an equivalent role in elevating cAMP, demonstrating therapeutic benefits in these diseases (Figure 4B. adapted and modified from Ham *et al*).[121–129] Hence, it would be informative to design ligands that would elevate cAMP at A_1R , A_{2AR} , and PDE10A and investigate the potential therapeutic benefits of a multi-target approach.

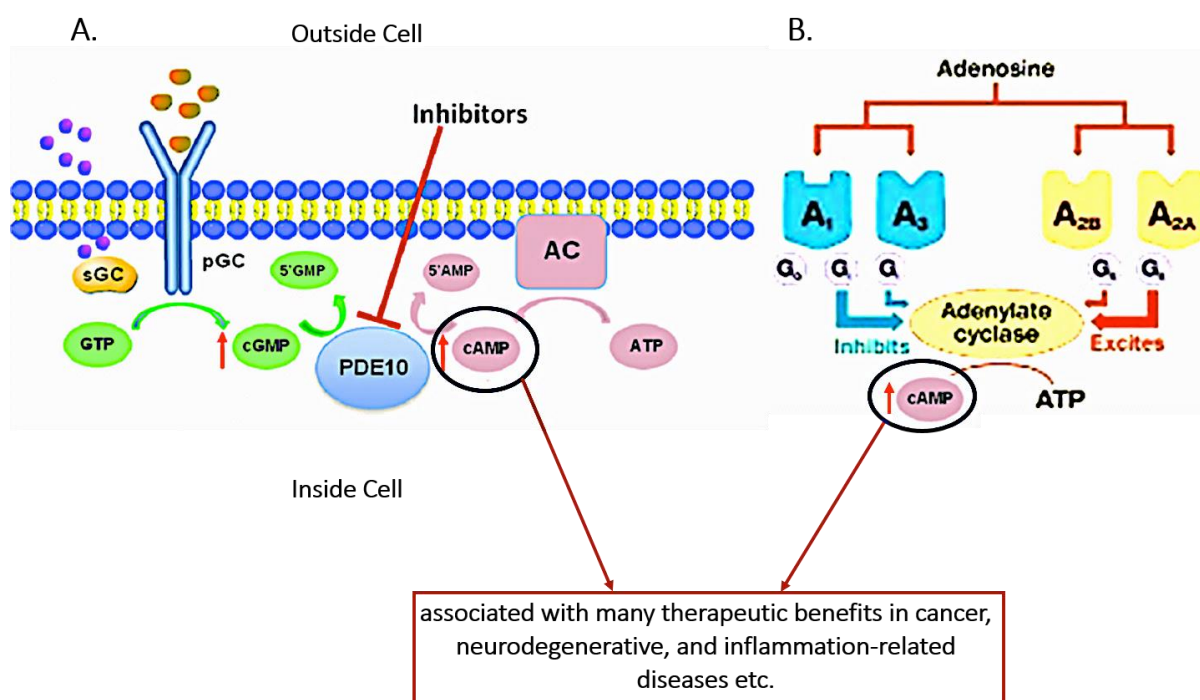


Figure 4. A. The inhibition of PDE10A results in the elevation of the concentration of cAMP which is suggested as an effective treatment in multi-factorial diseases such as CNS traumas, autoimmune disease, inflammatory diseases, and cancer [121–125] **B.** The antagonism of the A_1R and the agonism of the A_{2AR} , receptors plays an equivalent role in elevating cAMP, demonstrating therapeutic benefits in these diseases[127, 128]. Hence, designing ligands that elevate cAMP levels at the A_1R , A_{2AR} , and PDE10A targets would be interesting to be explored further for their therapeutic benefits in the aforementioned diseases.[96–98]

Dual PDE inhibition and A_{2AR} activation, *via* compound combinations, was synergistic in elevating cAMP, and was observed to inhibit proliferation of multiple myeloma and diffuse large B-cell lymphoma as well as induce apoptosis.[98] As such there is the possibility of targeting both the A_{2AR} and PDE10A, in particular, as an anti-proliferative strategy for the

treatment of diseases with up regulated proliferation, such as cancer. The A_{2A}R is expressed in e.g. both lung adenocarcinoma and squamous cell carcinomas.[130, 131] Similarly, PDE10A is overexpressed in lung adenocarcinoma, and its inhibition was found to suppress growth.[132]

To explore this approach, a computational method that combines ligand- and structure-based techniques is presented, which employs *in silico* target prediction[133] and docking[134] for the design of synthetically feasible multi-target ligands that bind to the A₁ and A_{2A} receptors and that also inhibit PDE10A. The method has been extended to consider functional effects and aims to identify chemical series that show agreement in both ligand- and structure-based predictions at A_{2A}R and PDE10A, as a sufficient dual-target combination to elevate cAMP.

In this work, known PDE10A inhibitors belonging to the chemical series were identified and shortlisted as A_{2A}R agonists *via* a structure-based approach, which consisted of docking and MD simulation. The MD simulation analysis also enabled the identification of a novel conformational descriptor characterizing A_{2A}R activation. This addresses one of the outstanding challenges in designing GPCR ligands that is predicting their functional effects. The compounds were then validated experimentally as A_{2A}R agonists in relevant biochemical assays, and subsequently tested for their anti-proliferative effects in lung carcinoma cell-lines.

1.2.5 Current computational approaches for the design of multi-target drugs

Rational drug design is widely practiced to find lead compounds, typically to generate leads or optimize candidate structures for a specific target. However, there is growing interest in finding multi-target ligands that exhibit a specific multi-target activity profile. Ligand- and structure-based computational approaches are employed in e.g. virtual screening (Figure 5) to generate starting structures for optimization.[57, 135, 144–148, 136–143] The use of structure- and ligand-based techniques in drug design, is however dependent on whether there is sufficient structural information of the drug targets/ligands of interest.

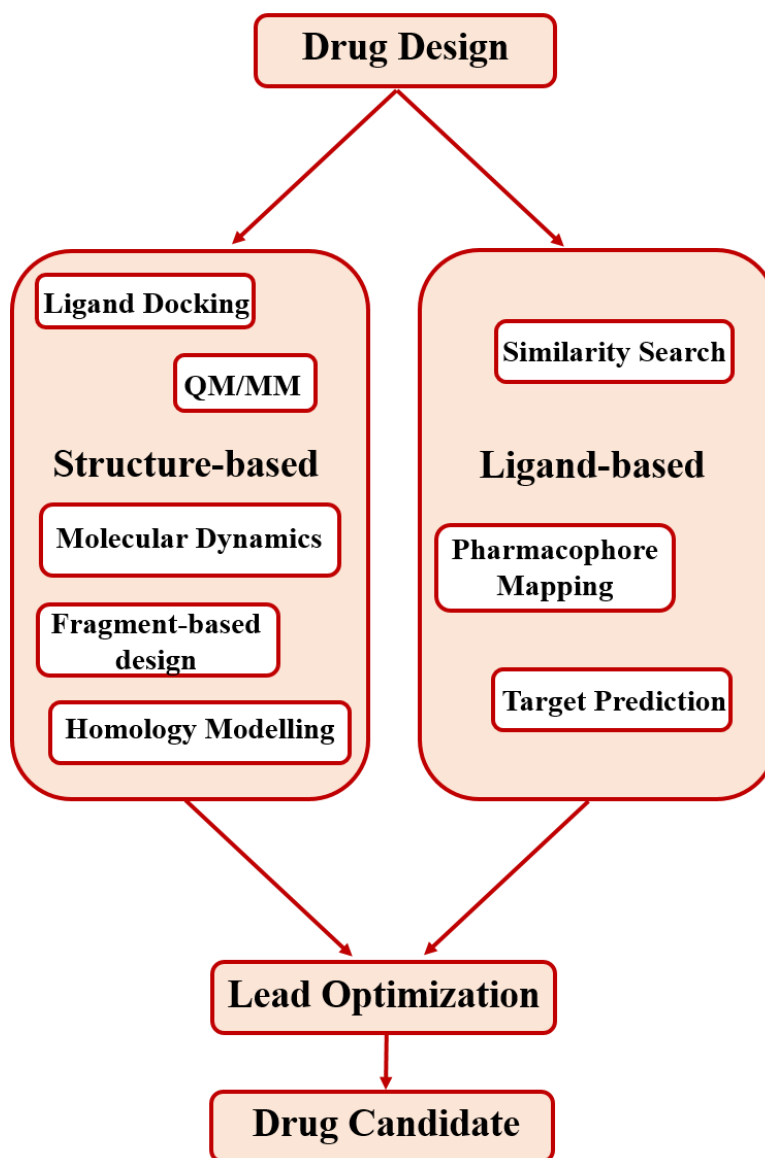


Figure 5. The use of structure- and ligand-based techniques in drug design is dependent on the available data. Where no structural information is available, this would lead to the sole use of ligand-based techniques, where known active compounds are used for the discovery of drug candidates based on similarity search, pharmacophore mapping or target prediction. Where structural information is available, e.g. a protein X-ray structure, then structure-based techniques may be employed. Structure-based techniques include docking, molecular dynamics, QM/MM, fragment-based design, and homology modelling etc. Both approaches can be employed in combination where data is available.

Ligand-based techniques

In the case where no structural information is available this would lead to the use of ligand-based techniques, where known active compounds are used for the discovery of similar and more suitable drug candidates (e.g. better pharmacokinetics or higher potency). The most commonly used ligand-based techniques are similarity search, pharmacophore mapping and target prediction. Similarity-based or fingerprint-based approaches select novel compounds on the basis of their chemical and physical similarity to known compounds with activity at the target of interest, implementing the theory that structurally similar molecules tend to have similar binding properties.[147] Quantitative structure–activity relationship (QSAR) models can be used to find a relationship between structures of molecules and their target response.[144] Pharmacophore mapping relies on a geometric molecular framework defining the fundamental features accounting for the bioactivity of a compound. Pharmacophore models generally employ key functional groups at defined locations for active ligands binding to the target of interest.[149] One of the pioneering pharmacophore modeling techniques was the active analog approach described by Marshall *et al.*[150]

Structure-based techniques

In cases where structural information is available for the targets of interest, then structure-based techniques may be employed either alone or combined with ligand-based techniques. The Protein Data Bank (PDB)[151] constitutes the biggest depository of protein structures determined by X-ray crystallography and NMR techniques and generally serves as the basis for most structure-based drug discovery projects. Structure-based techniques include docking, which predicts the preferred orientation of a compound to form a stable complex with a particular target of interest and a scoring scheme that can be used to rank structures, and binding solutions. In addition, molecular dynamics simulations describe the dynamics of proteins by simulating motion according to Newtonian mechanics.[144, 152] Furthermore, quantum and molecular mechanical (QM/MM) based methods may be employed to determine energetics of binding, reaction coordinates or to optimize the active site placement of ligands.[153] These methods are however very time consuming. Other structure-based methods include fragment-based design, which attempts to identify fragments (low molecular weight compounds) as weak binders to the target of interest that are subject to fragment growing, which adds additional fragments or functional groups to optimize interactions with the binding site. Fragment linking is an alternative approach to link fragments to adjacent sites of a protein target to increase

binding affinity. The objective is e.g. to generate leads with high affinity and selectivity.[154] In the case where the structure of the protein target of interest is not yet solved, homology modeling might be required. Homology modeling is used to build a model for the protein structure by the use of a structural template protein (homologous) of similar sequence.[144]

1.2.5.1 Virtual Screening in multi-target ligand design

Virtual screening (VS) is an inexpensive computational approach with an essential role in drug discovery assisting expensive experimental high throughput screening to optimize compound screening and improve hit discovery.[148] Both ligand-based (LB-) and structure-based (SB-) techniques can be implemented in VS. In fact, the application of SB-VS has been successful across various protein targets in drug discovery.[148] For example, potent inhibitors of Hsp90 (heat shock protein 90), a key therapeutic target in cancer, were discovered by applying SB-VS of several hundred thousands of compounds from the 1.62M compound rCat database [155] against Hsp90. Another example involves the discovery of novel series, 1-(N-substituted piperidin-4-yl) benzimidazolones, by using the homology model of the M1 acetylcholine receptor (mAChR) derived from the crystal structure of bovine rhodopsin.[148] In addition, the SB-VS protocol devised by Jaiteh *et al* led to the discovery of dual-target ligands at the A_{2A} adenosine receptor (A_{2A}R) and monoamine oxidase B (MAO-B), a polypharmacological profile relevant for Parkinson's disease. Similarly, compounds discovered by LB-VS and 3D-QSAR analysis, were successful in inhibiting MAO-A, MAO-B, AChE, and BuChE.[156] Hence, in many instances, the use of SB-VS and LB-VS have been successful in providing new avenues for drug discovery against multi-factorial diseases.[95]

Besides the separate application of LB-VS and SB-VS in drug discovery, the combination of both could also be implemented. For instance, Lepailleur *et al* discovered benzo[h]-[1,6] naphthyridine ligands as dual H₃R (Transcriptional elongation factor) antagonists/5-HT₄R (Serotonin 4 receptor) agonists by applying a combination of pharmacophore-based virtual screening, similarity based clustering methods, and molecular docking.[157] Binding experiments confirmed the affinities of the selected ligands towards H₃ and 5-HT₄ receptors. In addition, fragments that exhibited dual activity against β -secretase1 (BACE-1) and glycogen synthase kinase 3 β (GSK-3 β) protein targets, two structurally unrelated enzymes associated with the onset of Alzheimer's disease, were identified by a combined SB-VS and LB-VS protocol developed by Bottegoni *et al*. [136] Also Butini *et al* reported their combined SB-VS

and LB-VS of multi-target ligands at the dopamine, serotonin 5-HT_{1A} and 5-HT_{2A} receptors, which are key proteins in neurological disorders.[158]

Hence, VS is a cheaper and faster method than HTS, which could be used for screening large *in silico* libraries, and could be combined with HTS as a prior step. In addition, VS allows the investigation of totally novel compounds that have neither been synthesized nor purchased. However, it cannot completely replace HTS that experimentally tests the activities of hundreds to thousands of compounds against a particular protein target of interest, which is a fundamental step in drug discovery.

In this work, SB-VS is employed both separately and in combination with LB-VS. An important consideration in VS is to shortlist a list of candidates that are synthetically feasible, for this reason RECAP (Retrosynthetic Combinatorial Analysis Procedure) has been employed.[159, 160] Ligand-based approaches are used in combination with structure-based approaches, and in particular target prediction is employed to find compound series that show agreement with the predictions of multi-target ligands at the A₁R, A_{2A}R, and PDE10A *via* target prediction and docking. A comprehensive description of the ligand-based techniques employed is included in section 1.2.5.2. In addition, structure-based approaches, in particular molecular docking and molecular dynamics simulations, which are used in designing allosteric calpain-1 inhibitors and multi-targeted ligands that elevate intracellular cAMP are discussed extensively in sections 1.2.5.3 and 1.2.5.4 respectively.

1.2.5.2 LB-VS – target prediction

In general, ligand-based approaches are based on the principle of chemical similarity, where similar compounds are predicted to exhibit similar biological properties. Similarity searching for ligand-based *in silico* target prediction has been widely practiced. It is a popular approach that can predict whether candidate compounds are active against a particular target of interest[161] based on their molecular similarity to known bioactives reported in chemogenomic databases.[162, 163] Target prediction tools may be employed for this purpose. For instance, SwissTargetPrediction is a web server that implements 2D and 3D compound descriptors to predict whether test compounds are active (against a particular target) based on the similarity of the descriptors to those of known active compounds against the same protein target.[164] In addition, PIDGIN 1.0 (Prediction including Inactivity) is a tool that uses ECFP 4 circular Morgan fingerprints trained on ChEMBL actives and PubChem inactives to predict the activity of the query compounds, and it has been later updated to its newer version (PIDGIN

2.0).[133, 165] Another tool for target prediction, which accounts for the polypharmacological profile of the query compounds, is PolyPharmacology Browser (PPB). PPB is a web-based platform that searches for nearest neighbors to predict the multi-activity profile of small molecules using ten different fingerprints accounting for composition, substructures, molecular shape and pharmacophores.[166]

In this work, A₁R, A_{2A}R, and PDE10A were the protein targets of interest for the investigation of novel multi-target ligand design approaches. Target prediction using PIDGIN 1.0 was employed to find compound series that show agreement between target predictions and docking focusing on multi-target ligands at these targets. The series found by this method were synthesized and then experimentally validated as multi-target ligands in relevant biochemical assays.

PIDGIN 1.0 uses Naïve Bayes (NB) classifiers, which are probabilistic models implementing Bayes' theorem and assuming independence between the features used. This family of algorithms has been extensively used for bioactivity prediction, due to their rapid training and prediction times, as well as their insensitivity to noise.[167] Other example of a multi-class Naïve Bayes classification algorithm includes one that has been trained by Nidhi *et al* [168] on data composed of over 960 target proteins extracted from the 'World Of Molecular BioAcTivity' (WOMBAT).[169] Another example was developed by Koutsoukas *et al*, [170] for predicting structure activity relationships (SARs) of orphan compounds. This employed either a Laplacian-modified Naïve Bayes classifier or a Parzen-Rosenblatt Window (PRW) learning algorithm. The algorithm was trained on data from the ChEMBL14 database,[171] consisting of more than 155,000 ligand-protein pairs of 894 different human protein targets.

Given that these methods only consider the structure of the compound without taking into account that of the protein target, this may limit their predictivity. Hence, it would be best if ligand-based approaches are combined with structure-based techniques, which has been done here. This combination of techniques might provide a more complete picture of drug-target interactions, and it is expected to considerably increase confidence in the predictions being made, which would potentially yield a higher rate of success for rational drug design.[172]

1.2.5.3 SB-VS-Molecular Docking

Molecular docking is often used to predict the interactions between small molecules and protein targets, along with their optimum poses and conformations when fitting the protein-binding site as well as providing an estimate of the stability of the ligand–protein complexes formed.

Commonly used docking software include Glide[134], GOLD[173], DIVALI[174], DARWIN[175], CDOCKER[176], and AutoDock.[177] Docking is more efficient when the binding site of the protein target is experimentally determined by X-ray crystallography and available in the protein data bank (PDB). However, in the case where it is not known, information about the sites could be obtained by comparing them with a family of proteins of similar function, co-crystallized with other ligands. When nothing is known about the binding site, programs such as GRID[178, 179], POCKET[180], SurfNet[181, 182], PASS[183] and MMC[184] may be employed to predict potential binding pockets within proteins. In the case where docking is performed without knowledge of the binding site, this would be referred to as blind docking.[185]

Furthermore, it is important to find the most stable binding mode of a ligand by optimizing its geometry while docked to the protein target; hence, objective functions (e.g. molecular mechanics) are calculated and used to optimize docking. In addition, several sampling algorithms have been developed to estimate binding affinity using scoring functions, often used for ranking generated conformations or ranking series of binders. This helps avoid generating all the possible conformations that would be too computationally expensive when taking into account all the translational, rotational, and conformational degrees of freedom of both the ligand and protein.[35]

Various docking algorithms

Matching algorithms (MA)[186–188] use molecular shape to map ligands into the active site of a protein taking into account chemical information and shape features. For this purpose, calculations are performed that generate a distance matrix, which aids in the search for ligand conformations that lead to the best match between pharmacophores representing the protein and the ligand atoms. In addition, chemical properties such as hydrogen-bond donors and acceptors are generally considered in these matching algorithms. These algorithms are used since they are fast, they enrich active compounds from large libraries,[189] and they are implemented in programs such as DOCK[190], FLOG[191], LibDock[192] and SANDOCK.[193]

Incremental construction (IC) methods[194–196] place the ligand into the binding site in a fragmental and incremental manner, where the ligand is fragmented into many parts often by breaking rotatable bonds. Subsequently one of these fragments is chosen to be docked first into the active site, usually the largest fragment or one with a major functional role in contributing

in the interactions with the protein (e.g. a warhead structure). Then the remaining fragments are added incrementally, generating various orientations and conformations. This requires some flexibility of the docked ligands, until their optimum fit is obtained in the active site. These methods may be used in programs such as Glide,[134] DOCK 4.0,[197] FlexX,[194] Hammerhead,[198] SLIDE[199] and eHiTS[200].

Multiple Copy Simultaneous Search (MCSS)[201, 202] and LUDI[203] are also fragment-based methods that are used for the de novo design of ligands. They rely on modifying known ligands to improve their binding to the protein target. Around 1,000 to 5,000 copies of functional groups are generated by the MCSS method and randomly positioned into the binding site, while concurrently exposed to energy minimization and/or quenched molecular dynamics within the force field of the protein. The copies are only allowed to interact with the proteins but not among themselves. Then, based on the interaction energies, a group of energetically favorable binding sites and orientations for the functional groups is identified, enabling mapping the whole binding site using different functional groups. By linking those different functional groups, new candidate compounds may be designed that optimally fit the mapped binding site (at least, optimum in terms of the fitness function employed).

LUDI is a computer program which is focused on the interaction sites between the ligand and protein that are treated as distinct locations in space suitable for hydrogen bonding or hydrophobic interactions.[203] These interaction sites may be generated either by using rules or searching the database. Fragments are then placed into the interaction site and assessed by their distance to this site; by joining all the fitted fragments this would form a single compound.

Monte Carlo (MC) algorithms belong to stochastic methods, whereby it is possible to generate many ligand poses[204, 205] *via* bond rotation, rigid-body translation or rotation. Based on their interactions (and internal) energies, the ligand conformations are assessed, and if they pass, are modified to produce subsequent conformations. This is performed iteratively until a pre-defined quantity of conformations is assembled, allowing ligands to overcome energy barriers on the potential energy surface. Monte Carlo methods may be used in programs such as Glide, AutoDock (earlier version)[206], ICM[207], QXP[208] and Affinity (Accelrys)[209].

Genetic algorithms (GAs) are another class of well-known stochastic methods.[177, 210, 211] The idea for GAs originates from Darwin's theory of evolution. In computer-based implementations, binary strings, which can be termed genes, are used to encode the degrees of

freedom of the ligands. The three main genetic operators in GAs are mutation, crossover, and selection. Random changes to the genes are made during mutations. Genes are exchanged between chromosomes during crossover (this is in fact the most effective optimization process). The fittest combinations are generally selected for further optimization by using a scoring function. New ligand structures result when genetic operators modify the genes. Scoring functions are used to assess the new structures, and those that are above a selection threshold may be used for the next generation. GAs are used in programs such as AutoDock[177], GOLD[173], DIVALI[174] and DARWIN[175]. There are differences in the selection algorithms employed e.g. using Darwinian (Gold) or Lamarkian (AutoDock) evolution.

In this work, Glide is employed for protein-ligand docking in structure-based approaches for the design of allosteric calpain-1 inhibitors and multi-target ligands at A₁R, A_{2A}R, and PDE10A. Glide's docking algorithm is represented as a docking "funnel" in Figure 6. It utilizes a hierarchical search protocol and estimates a full systematic search over ligand positions, orientations, and conformations in the receptor site. Then minimization of the poses selected by the initial screening is performed using the OPLS-AA force field in combination with a distance-dependent dielectric model. Subsequently, the lowest-energy poses are subject to a Monte Carlo procedure to examine nearby torsional minima. Finally, the total Coulomb-van der Waals energy (E_{cvdw}), with the Coulomb energy screened by a distance-dependent dielectric constant determines the final ligand pose that is selected. The scoring function, GlideScore, which is used to calculate binding affinity, is an extension of an empirically based scoring function, Chem-Score function of Eldridge *et al.*[134]

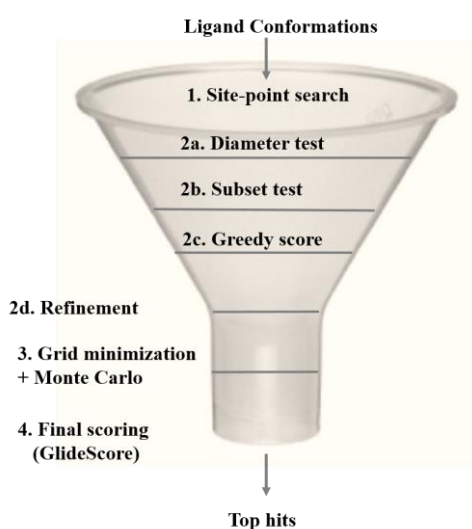


Figure 6. Glide docking "funnel" shows the Glide docking hierarchy. [134]

Scoring functions

Given that binding of ligands to their targets depends on their molecular interactions, physicochemical properties, as well as their conformations, scoring functions including these properties have been used in docking. Scoring functions are used to select the right poses of the ligands in a realistic time frame, where many assumptions and simplifications are made.[212] They are classified into three categories: force-field-based, empirical and knowledge-based scoring functions.[212]

Force-field-based scoring functions,[213–215] are used to evaluate the binding energy by calculating the sum of steric, electrostatic and van der Waals interactions. For electrostatic interactions, a distance dependent dielectric term is used with a cutoff distance for inclusion of electrostatics. A Lennard-Jones potential is used to describe the van der Waals interactions. Implemented force-field-based scoring functions have limitations e.g. the cut-off distance employed to deal with the potentially large number of electrostatic and van der Waals interactions decreases the accuracy of computing interaction energies over longer distances. Furthermore, entropy contributions, polarizability and solvation effects could be accounted for in extensions of force-field-based scoring functions. These extensions are offered in software such as DOCK[216] GOLD and AutoDock, with differences in implementation in each program. Moreover, refinement of the docking results with these functions can be combined with other techniques, such as linear interaction energy[217] and free-energy perturbation methods (FEP),[218] which could potentially improve the accuracy of binding energy prediction.

In the empirical scoring functions[219–223] the binding energy is divided into a number of components: hydrogen bonding, ionic interactions, hydrophobic effects and binding entropy. A final score is obtained by adding each energy component multiplied by a coefficient that is derived from regression analysis fitted to known ligand-protein complexes with reported binding affinities.

Despite the simplicity of the energy terms that are evaluated in the empirical scoring functions, it is not clear whether their applicability domain is suited for ligand-protein complexes that are not included in the training set. The applicability domain is the region in the ligand-protein chemical activity space where the scoring functions would be reliable. Moreover, various software packages treat each term in the empirical scoring functions differently. Examples of programs using empirical scoring functions include LUDI, PLP,[224] and ChemScore.[225]

In knowledge-based scoring functions,[226–231] crystal structures of ligand-protein complexes are used to perform statistical analysis of the ligand-protein distances and/or the frequencies of their interatomic contacts. It is assumed in these analysis that the higher the frequencies of the protein-ligand interactions the more they are favorable. The distributions of frequencies are transformed into pairwise atom-type potentials, where a score is computed according to a cut-off that considers favored interactions and penalizes repulsive ones.

There are many advantages to knowledge-based functions, such as their simplicity and speed, which allows their use for screening large compound databases. Also these functions may be used in modelling uncommon interactions (if data were available and included in the scoring function generation) such as cation- π and halogen- π -aromatic interactions, which might not be accounted for in empirical approaches. Some interactions, in particular those involving metals or halogens, are still not well represented in training sets that might involve insufficient number of crystal structures, which limits their applicability domains. Examples of knowledge-based functions include PMF,[226] DrugScore,[232] SMOG[233] and Bleep[227] that vary in their distance cut-offs, size of training sets, the form of their energy functions, definition of atom types, and training data etc.

Consensus scoring[234] has been used to evaluate docking and scoring by combining the results of several scoring functions. When the pose of a candidate ligand scores well across different scoring functions, it could be more confidently predicted to be a potential binder. This approach has been seen to improve enrichment and prediction of bound conformations and poses in VS.[235] However, accurate binding energy prediction is a severe challenge. There are some (named) scoring functions that are already a combination of other scoring functions e.g. CScore[236] is an example that combines DOCK, ChemScore, PMF, GOLD, and FlexX scoring functions.

Scoring functions in general share a common problem in affinity prediction can be attributed to limitations in considering solvation effects, entropy, kinetics, heat capacity and the generally inaccurate nature of empirical scoring. An approach to move towards addressing this could be the use of physics-based scoring functions such as MM-PB/SA and MM-GB/SA (MM stands for molecular mechanics, PB and GB for Poisson-Boltzmann and Generalized Born, respectively, SA for solvent-accessible surface area). Some studies suggested improved results obtained with MM-PB/SA[237, 238] or MM-GB/SA.[239] Nonetheless, it was reported that the GB/SA model poorly estimated protein desolvation in certain systems. In contrast, better

results were obtained when WaterMap was incorporated into the MM-GB/SA method instead of GB/SA protein desolvation.[240] Many comparative studies for various methods of affinity evaluation for protein-ligand complexes have been reported in the literature. One suggestion was that PDLD/S-LRA/ β (protein dipoles Langevin dipoles linear response approximation) might be a good option for use in the final stages of VS. As for PB/SA, it might not provide an accurate estimate of binding energies because of its erroneous estimation of entropies and its mishandling of electrostatic energies.[241] Recent work on Deep Learning and artificial intelligence (A.I.) methodologies claim to improve scoring performance.[242, 243]

In this work, GlideScore is the scoring function used in shortlisting ligands from the structure-based design approaches. It is an extension and a modification of ChemScore, an empirical scoring function.[134] The equations of ChemScore and GlideScore functions are discussed further in Appendix A.

Two forms of GlideScore may be employed, GlideScore 2.5 SP, Standard-Precision Glide and GlideScore 2.5 XP Extra-Precision Glide.[134] The GlideScore 2.5 SP is a “softer”, more lenient function, which predicts ligands to bind to the protein target of interest with reasonable affinities despite the imperfections in their poses. Thus, this GlideScore version would be more convenient for screening databases to minimize as much as possible false negatives. In contrast, GlideScore 2.5 XP is a stricter function that applies more stringent penalties for poses violating e.g. atom pair interactions. This version would then minimize false positives, which could be useful for lead optimization or for the selection of the final top candidates by VS.[134] In this work, the choice of SP or XP parameters in docking has been deduced from experiments performed on known actives and inactives for each of the protein targets considered, allowing for the maximum separation between their docking score distributions.

Docking methodologies

Rigid ligand and rigid receptor docking is one type of docking, where both the ligand and receptor are kept rigid, by limiting their space to three translational and three rotational degrees of freedom. This method can account for the ligand flexibility by a pre-computed set of ligand conformations, or by enabling some degree of atom–atom overlap between the protein and ligand. It has been used in programs such as DOCK, FLOG, Glide and some protein-protein docking programs, such as FTDOCK.[244]

The first automated procedure for ligand-protein docking was DOCK, and it has been constantly improved. In this program, the ligand and receptor are treated as sets of spheres that

could be overlaid using a clique detection procedure.[245] And the scoring of the ligand-protein complexes is done according to some criteria such as pharmacophore similarity, chemical complementarity, or steric fit, where geometrical and chemical matching algorithms are implemented.

Flexible ligand and rigid receptor docking is another type of docking that could be used for ligand-receptor systems behaving within the induced fit paradigm. Usually it would be better to account for the flexibility of the ligand and the receptor, since both change conformation on association, and in particular in the cases where the receptor is highly flexible. However, sometimes a compromise between accuracy and computational time has to be made where it would be more feasible to keep the receptor rigid and the ligand more flexible during docking. This method is implemented in many docking programs such as AutoDock, FlexX, and Glide etc.[246, 247]

Flexible ligand and flexible receptor docking is a third type of docking. As suggested by the literature, the flexibility of proteins is affected by ligand binding (induced fit);[248] however, it would be very challenging to account for receptor flexibility in docking alone. MD simulations could better account for protein flexibility by modeling most of the possible degrees of freedom of ligand-receptor complexes. Yet, sometimes due to the high computational expense of MD and inadequate sampling, this method might not be suitable for screening large chemical databases.

Furthermore, various theoretical models - conformer selection, conformational induction, and historic induced fit, have been proposed to account for flexible ligand-protein binding. When a ligand binds selectively to a favorable protein conformation over a number of possible protein conformations, this is referred to as conformer selection. As in the case of conformational induction, the ligand transforms the protein's conformation upon binding from what was in the unbound state.[174]

“Soft-docking” is a method used to account for receptor flexibility[249] by allowing for some atom-atom overlap between the ligand and the receptor *via* modifying the van der Waals repulsion energy term in the scoring function used. Examples where this method is implemented include the LJ 8-4 potential in GOLD and the smooth potential in AutoDock 3.0. Given that the receptor coordinates are fixed in this method, and it simply relies on modifying the van der Waals parameters, this optimizes its computational efficiency, though it might result in inadequate flexibility.

Using rotamer libraries is another method for modeling receptor flexibility.[250, 251] These libraries consist of side-chain conformations derived from statistical analysis of structural experimental data. Employing this method offers relatively quick sampling, and it avoids minimization barriers. Coupled with biased probability methodology,[207] and Monte Carlo search of ligand conformations, rotamer libraries are used in the ICM (Internal Coordinates Mechanics) program.[207]

Additionally, in order to account for side chain flexibility, AutoDock 4 adopts a simultaneous sampling method,[252] where users can select several side chains of the receptor to be sampled. As for the rest of the receptor it is dealt with as rigid during sampling with an energy grid map. Using a combination of protein conformations, utilizing the theory of conformer selection, would be another way to account for receptor flexibility.[152, 253] Ligands may be separately docked to each of the rigid protein conformations, in order to merge the results depending on the method chosen.[254] This method was first used in DOCK, where it produces an averaged potential energy grid for a combination of protein conformations.[152] Also in other programs such as FlexE several crystal structures of a particular protein are used, merging analogous parts and marking the dissimilar ones as different parts to be considered as alternatives. Protein conformations are sampled in a combinatorial manner during the incremental construction of a ligand, where the best scoring is selected based on the comparison between the ligand and its alternative.

Hybrid docking is another approach used for modelling receptor flexibility. This method is implemented in Glide, and uses a set of hierarchical filters in the search for possible ligand poses and orientations within the receptor-binding site. As for ligand flexibility it may be accounted for by performing a thorough search of the ligand torsion energies, based on which ligand conformations are selected with a soft potential for docking into the receptor binding sites. Subsequently, to further model receptor flexibility rotamer exploration may be employed. For instance, IFREDA[254] accounts for receptor flexibility by employing a hybrid method that combines soft potentials and multiple receptor conformations. As for other programs like QXP and Affinity, a Monte Carlo search of ligand conformations may be done accompanied by a minimization step, where the user defines parts of the protein that are allowed to move to avoid clashes between the ligand and receptor. In addition, flexibility may be incorporated with SLIDE by removing clashes through directed single bond rotation of either the side chains of the protein or the ligand. In order to model induced-fit complementarities between the ligand and protein, an optimization approach based on the mean-field theory may be applied.

All of the methods mentioned either deal with the side chain or the full receptor flexibility. However, sometimes loops constituting the active sites of several enzyme families such as Bromodomains,[255, 256] contribute a major role in ligand binding, and undergo dramatic conformational changes while keeping the other parts of the receptor unchanged. In this case, methods that account for side chain and full receptor flexibility fail to deal with loop flexibility. Local Move Monte Carlo (LMMC) loop sampling method could be used in this case. This method relies on sampling ligand conformations within loop-containing active sites. It is the torsion angles resulting from side chain movements as well as the motion of the loop backbones that generates the allowed loop conformations. A grid-based force field that represents the protein environment and solvation effects was developed to minimize the computational costs for the evaluation of energy in these calculations.[257, 258]

In this work, the flexible ligand and rigid receptor docking methodology in Glide is employed for the structure-based design of allosteric calpain-1 inhibitors and multi-target ligands at A₁R, A_{2A}R, and PDE10A. This docking methodology has been selected since all the protein targets of interest are fairly rigid as assessed by thermal stabilities (B factors)[259] in Glide.

1.2.5.4 SB-VS-Molecular Dynamics Simulations

Molecular dynamics (MD) is extensively used in molecular modelling. MD can account for the flexibility of the docked ligand-protein structures by moving each atom separately in the field of the other atoms in order to rationalize or even predict experimental behavior of the system. [260–262] MD is often used as a sampling technique to explore as many configurations as possible, which allows for the identification of low energy configurations or the calculation of the system's equilibrium or dynamic properties. It may be used to generate trajectories to study the system's kinetics e.g. folding of a protein.[263] Programs such AMBER[264] and Desmond[265] are available to perform MD simulations.

First, a structure of the ligand-protein complex is prepared, then mathematical equations (Equations 1-3)[263] are used to estimate the forces acting upon each atom of the system (molecular mechanics (MM)). The trajectories of the atoms are determined by numerically solving Newton's equations of motion.

MD is an efficient method, which although relatively fast (compared to QM), enables consideration of the physics of biomolecular systems with an approximation that neglects the quantum effects. This is the Born-Oppenheimer approximation, where each particle in the system, whether a single atom or a rigid set of atoms, is treated as a point mass i.e. only the

nuclear displacements need to be considered, and not the quantum mechanical effects of the motions of the electrons.[266]

The relationship between the velocity of an atom i , and its momentum, p_i , with a mass m_i , and a position indicated by the 3-dimensional vector r_i , is illustrated in Equation 1.

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} \text{ (Equation 1)}$$

Equation 2 shows the net force, F_i , exerted on the atom i by the remainder of the system as the negative gradient of the potential energy function in relation to its position.

$$F_i = -\frac{dV}{dr_i} \text{ (Equation 2)}$$

Equation 3 is the Newtonian equation of motion for atom i .

$$\frac{dp_i}{dt} = F_i \text{ (Equation 3)}$$

In an MD system, the position and the velocity of each particle is recorded and Equations 1-3 are iteratively and simultaneously solved over short time steps, whereby the system's deterministic trajectory is computed. Given the discrete nature of the calculations, the Verlet algorithm[267] is employed as a correction. See Appendix B.

The process is repeated iteratively as the simulation time progresses. An important consideration is that the time step has to be substantially shorter than the motions of the highest frequencies, which are the C-H stretches. Otherwise simulations might be inaccurate and deviations from the standard bond lengths might occur resulting in high forces and velocities that could lead to the “explosion” of the system. The standard time step is 1 fs, in order to increase the time, constraint algorithms are usually applied. The SHAKE algorithm[268] is often implemented and constrains the X-H bond lengths, eliminating the fastest motions, which allows increased time steps to 1.4 fs. The LINCS[269] and M-SHAKE[270] algorithms may further extend the time step to 2 fs by constraining all bond lengths. It has been reported that it would be possible to increase the time step further to 5 fs for instance, in the case where there are no explicit hydrogen atoms in a system using the LINCS algorithm.[271]

An important consideration in any MD is the choice of the system ensemble[272] that has to be computed taking into account the nature of the simulation that is being performed. The factors constituting the simulation are the conservation of each of: matter (N), volume (V) or pressure (P), and temperature (T) or Energy (E). A canonical ensemble (NVT) is used in the simulations. In general, the ensemble used is the Isothermal-Isobaric (NPT), since biological

systems are normally maintained at constant pressure conditions with little change in temperature. In both types of ensembles a thermostat e.g. Berendsen[273] or Nose Hoover, is used to keep a constant temperature by adding or removing energy from an isothermal bath. In the case of the NPT ensemble, a barostat[274] is required e.g. Berendsen or Parrinello-Rahman to sustain a constant pressure.

Periodic boundary conditions (PBC)[275] are employed in most biomolecular simulations in order to avoid artefacts at the edge of the simulation box. By applying PBC, leaving one face of the system (e.g. in a cubic system) would take the particle to the “opposite” side of the system, which can e.g. be implemented as a cube, octahedron etc. PBC better accounts for an ‘infinite’ system, though it also exhibits artefacts arising from the interactions of the molecules with themselves in small systems, or interacting twice (in both sides of the box for example) with other molecules in somewhat larger systems. Hence, to ensure the best performance of PBC, the sizes of the system should be substantially larger than the interaction cut-off distances.

Reproducing realistic motions of atoms requires all the necessary potential functions and parameters, optimized to be self-consistent, and this is termed a force field.[263] A force field generally employs a set of well-defined equations for bonding and non-bonding interactions, which contain harmonic potentials for bond lengths, bond angles, torsion angles and associated force constants, pair-wise calculated van der Waals and Coulombic functions with associated parameters and electrostatic terms.[276] Depending on the equations used and the parameters incorporated into them, which are determined for the systems of interest, this would determine which force field is selected for a particular problem. Commonly used force fields in Biological MD simulations are AMBER, CHARMM,[262] GROMOS,[277] and OPLS3,[278] which are parameterized differently.[279] In some force fields, the parameters are based on experimental data and quantum mechanics calculations, with assumptions and approximations implicit in the force field equations and MD simulations, which often result in inaccurate “macroscopic” results. In other force fields each parameter could be virtually meaningless by its own, but rather all together they would result in simulations that are in agreement with experimental observations. In general, the latter types of force fields perform better in the systems that they are parameterized to work in, but poorly in systems that fall outside the scope of parameterization, which is often attributed to over-fitting. The majority of force fields lie between the two types by primarily using parameters derived from theory and experiments, which are then subject to some modifications in order to improve the simulations.[276]

There are many challenges in terms of computational costs that are encountered in large-scale simulations with non-bonding interactions, since every atom has to interact with every another atom. Also in the case where PBC is used, atoms might interact with themselves, which might result in infinite interactions. One of the easiest ways to solve these issues is to use a cut-off distance, beyond which no interactions would be calculated. For instance, a cut-off of 1.0-1.5 nm is suitable for van der Waals interactions but not the long-range electrostatic interactions, where large errors might result, intensified by the abrupt change at the cut-off distance from full interaction to no interaction.[276] The commonly used techniques to account for this are: the Reaction Field (RF),[280] which merges the electrostatic interactions into a simple field beyond the cut-off; Ewald summation that uses Fourier transforms to sum up electrostatic interactions to infinity over PBC; Particle Mesh Ewald (PME),[281, 282] where the summation is done upon a mesh, which is a faster but somewhat less accurate form of the Ewald summation.

Finally, the use of implicit atoms is one of the fundamental properties characterizing force fields. A dielectric constant and an adequately parameterized force field may be employed to approximate solvent molecules in order for them to be completely removed from simulations. Though this might substantially decrease the accuracy of the simulations, an advantage is that the computational cost is massively reduced. For instance, the implicit hydrogen atoms are used in united atom force fields such as GROMOS[277] to reduce the computational load. Given that the X-H bond length is short, since the van der Waals radius of the hydrogen gets shortened by the atom it is bonded to (in the case of X=C for example) it would be possible to approximate the $-CH_2$ and the $-CH_3$ to a sphere that is slightly bigger than carbon itself.

In this work, Desmond[283] with the default force field OPLS3 was employed to perform MD simulations. The choice of this force field is that it has been reported in the literature to include much more reference data and associated parameter types in comparison to other force fields often used for small molecules (e.g. MMFF and OPLS_2005).[278] In addition, across different validation studies, OPLS3 appeared to exhibit higher accuracy when assessing different conformational and solvation properties of small molecules. A detailed description of the functional form of the OPLS3 force field is included in Appendix C.

However, one of the limitations of MD is its high computational demand for simulation of large systems (although clearly not as high as QM). Implementations for the simulation of large biological systems generally require many processors and a significant time to be completed.

In addition, another disadvantage is the over-simplification of the force fields that are used to replicate the quantum-mechanical reality of the simulated systems, in particular where electronic effects (e.g. polarization (not generally present in most current force fields), aromaticity, dispersion) are important. These can be significant e.g. when transition metal atoms are involved in binding. Another limitation is the small-step progression, which makes it difficult to surpass conformational barriers of high-energy in a realistic time-scale that might result in inadequate sampling. Nonetheless, MD is an effective method in local optimization, and for this reason it is beneficial to carry out MD simulations after selecting the best conformation of the ligand/protein complex. Hence, an effective approach is to combine molecular docking with molecular dynamics simulations to predict key molecular interactions and dynamics upon ligand binding to protein targets. This can then be associated with potential biological effects.[263] The described approach is implemented in this work, where docking is followed by MD simulation analysis to find a conformational descriptor characterizing receptor activation.

Given that e.g. the A_{2A}R is a GPCR, which is a transmembrane protein, then it is essential to mimic its natural environment in order to correctly predict its dynamics in MD simulations. The thickness of the cell membrane is approximately 7 nm, and it consists mostly of phospholipids with other molecules such as glycolipids and steroids. It is amphipathic by nature i.e. extremely apolar from the inside (mainly hydrocarbon) and polar from the outside. The homogenous bilayer made up of an ensemble of a single phospholipid is commonly used to simulate the membrane. It is capable of mimicking many of the physical properties of the bi-layer despite being less structurally diverse than the real cell membrane. The use of the homogenous bilayer of a single phospholipid avoids artefacts resulting from poor mixing or incorrect component ratios, and simplified molecular topologies. 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC),[284] 1-palmitoyl-2-elaidoyl-phosphatidylcholine (PEPC),[285] 1,2-dimyristoyl-phosphatidylcholine (DMPC),[286] and 1,2-dimyristoyl-sn-glycero-3-phosphocholine and dipalmitoylphosphatidylcholine (DPPC)[271] are the commonly used phospholipid bilayers in the simulations of GPCRs. It has been reported in the literature that the use of POPC allows maximum mobility and flexibility of the receptor,[287, 288]. In this work (simulations of the A_{2A}R) POPC has been selected to construct the membrane used in the simulations.

1.2.5.1 Examples of successful applications of molecular docking and MD simulations in drug discovery

Molecular docking has been extensively used in SB-VS and has been successful in finding novel bioactives[289, 290] For instance, when docking was combined with other computational techniques and experiments, it has assisted in the analysis of drug metabolism by increasing understanding of the cytochrome P450 system.[291–293] Furthermore, the structure based design performed by Boehm *et al* using LUDI and CATALYST resulted in several novel and potent DNA gyrase inhibitors, where HTS proved ineffective.[294] In addition, Doman *et al* performed a comparative study for SB-VS and HTS,[295] where both were applied to screen protein tyrosine phosphatase-1B (PTP-1B) inhibitors. The results of the study highlight the success of SB-VS, where the hit rate was enriched by 1700-fold *via* docking as compared to random screening. It has been noted however, that the hits are more diverse in the VS and HTS, which suggests that it might be beneficial to combine both for lead discovery.

Furthermore, docking and MD simulations using the Autodock Vina program were performed for synthesized donepezil-indolyl hybrids[156] and donepezil-pyridyl hybrids[296] against the PDB crystal structures of four enzymes (AChE/BuChE/MAO-A/MAO-B). MD simulations identified the most promising donepezil hybrid[297] as a good starting point for analog modification for the design of novel multi-target ligands for the treatment of AD.

Hence, docking and molecular dynamics simulations are key tools in drug discovery, and given all their successful implementations in pharmaceutical research, they have been employed here in structure-based ligand design approaches. Docking is used in the structure-based design of allosteric calpain-1 inhibitors, also it has been combined with molecular dynamics simulations to design and rationalize the functional effects of multi-target ligands that elevate intracellular cAMP.

1.2.5.2 Computational approaches for lead optimization and prediction of physicochemical properties for compounds

QSAR modelling

In order to optimize the physico-chemical and biological properties of the identified lead compounds, and to efficiently minimize experimental work, quantitative structure–activity relationship (QSAR) modeling may be used. This approach is used to predict the effects of changes in molecular structure upon the physico-chemical properties of compounds. For this

purpose, representative molecular descriptors are selected as molecular features that are related to bioactivity, and which are, based on other studies, shown to be useful in QSAR modeling, and also have been used to assess structural similarity or diversity.[298, 299] 2D-QSAR methods tend to require less intensive calculations and therefore these are used as preliminary filters to screen compounds for later stages of drug development. In 3D-analyses, the usual approach is to construct pharmacophores from the geometric disposition of key functional and binding groups and are used to define the functional associations between the 3D-molecular determinants and bioactivity.[299]

The results of QSAR analysis may be used to rationalize favorable molecular interactions between the protein and the compounds of interest. For instance, the design of tacrine- and donepezil-like multi-target ligands for the treatment of Alzheimer's diseases constitutes an example of the use of 3D-QSAR approaches which are successful[300, 301] However, care is required in the use of QSAR modeling, as there are drawbacks associated with, in particular, over-fitting the data. False connections can be drawn when there are many experimental errors in the biological data (not uncommon in large databases)[62, 299], or there are few examples, or very similar bioactive compounds in the training set. Hence it is vital that a QSAR analysis is associated with an applicability domain assessment.[302]

Prediction of CNS permeability

For drugs that act centrally (CNS), it is important to understand drug partitioning across the blood–brain barrier (BBB). This is particularly relevant in this work, as all of the ligands designed (targeting PDE10A, A₁ and A_{2A} receptors, and calpain-1 proteins) are required to act centrally. Hence, it would be essential to understand CNS permeability of the studied ligands.

CNS permeability is determined by several factors such as the ability to penetrate the BBB (between the blood capillaries and brain tissue, active and passive transport and diffusion across membranes), the extent of their distribution in the brain, and their activities at the targets. Taking all of these aspects into account, this increases the complexity of CNS drug discovery.[303] One important parameter to be considered is the drug unbound brain concentration ($C_{u,b}$), which is determined by the drug concentration at the site of the target, and how it affects the *in vivo* drug efficacy. Another important experimental parameter is the receptor occupancy, which is indicative of the level of target engagement by the drug. As for the total brain concentration (C_b), this indicates the level of nonspecific binding of the drug in the brain.[304, 305] It would be possible then to predict the efficacy of CNS drugs by drawing

quantitative structure-exposure relationships that relate the experimental parameters of brain exposure to the molecular parameters characterizing drugs.[305, 306]

CNS drug efficacy and its penetration in the brain are favored by optimizing drug physicochemical properties, in particular lipophilicity, hydrogen bonding, aqueous solubility, pKa, and molecular weight. Transport of CNS drugs is facilitated through the BBB by moderate lipophilicity at a physiological pH of 7.4, where cLogP and cLogD are both in the range of 2–5.[307] Higher lipophilicity may be associated with an increase in plasma protein binding leading to a decrease in drug solubility in the plasma, and increased metabolic and toxicity risks.[308]

Unbound brain concentration ($C_{u,b}$) of CNS drugs is a major contributor to their *in vivo* drug efficacy, which is controlled by hydrogen bonding parameters such as hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA) counts.[309] Accordingly, parameter ranges have been determined from analysis of available data that appear to optimize exposure in the CNS. These are to decrease the count of the HBD and HBA parameters to $HBD < 3$ and $HBA < 7$. [310] In addition, the unbound brain concentrations of CNS drugs may increase with moderate lipophilicity ($cLogP < 4$) and a topological polar surface area (tPSA) range of 40–80 Å². Also, it appears that CNS drugs may exhibit a lower safety risk when the aqueous solubility is more than 100 μM.[311]

Programs such as QikProp in the Schrödinger[312] and FAF-Drugs3[313] software may be used to predict CNS permeability. It is still a challenging task to design CNS drugs with optimized physicochemical properties enabling efficient brain exposure.[307, 314, 315] In order to achieve more efficient penetration into the brain and gain enhanced efficacy, the computational predictions require further development including further exploration of the physicochemical properties space for available data on CNS drugs. In this work, FAF-Drug3 is employed to predict the CNS permeability of the compounds shortlisted by the various design approaches.

1.3 Aims of this work

In this work, single- and multi-target ligand design approaches for multi-factorial diseases were explored.

In chapter 2, calpain-1, which constitutes a promising therapeutic target in a number of disease areas including cardiovascular, neurodegenerative and ischaemic disease was explored. The discovery of selective calpain-1 inhibitors has been extremely challenging due to the similarity of its active site to a wide range of cysteine proteases[316] In this chapter, a structure-based virtual screening protocol, which employs docking, was devised to design allosteric inhibitors in an attempt to address the issue of selectivity of inhibitors against this enzyme.[317] Previously, selective calpain-1 inhibitors, such as PD150606[18], which included a specific α -mercaptoacrylic acid sub-structure (the chemical structure is depicted in Figure 7), were reported to bind to the penta-EF hand calcium binding domain, PEF(S) as well as the active site domain.[18] Although these are selective to calpain-1 over other cysteine proteases, their mode of action has remained elusive.[318] The structure-based virtual screening protocol reported here is a novel approach for the discovery of PEF(S) binders that populate a novel chemical space. This approach aims to elucidate an allosteric mechanism of action, which may offer improved selectivity and a reduced side-effects profile.

In chapter 3, a multi-target approach was investigated in order to find compounds, that in a targeted fashion for a specific disease state, elevate intracellular cAMP. Specifically, the adenosine receptor (A_1R and $A_{2A}R$) targeted elevation of cAMP has already been shown to be beneficial for many multi-factorial diseases such as in CNS trauma, autoimmune diseases, inflammatory diseases, and cancer.[96–98, 115, 116] Similarly, the elevation of intracellular cAMP concentrations upon PDE10A inhibition was suggested to be effective in the treatment of these diseases.[98, 117–121] Designing compounds that target this novel combination of G protein-coupled receptors (GPCRs) and an enzyme has not been previously exploited. A computational method that combines *in silico* target prediction and docking for the design of synthetically feasible multi-target ligands, which bind to the A_1 and A_{2A} receptors and inhibit the phosphodiesterase 10A (PDE10A) enzyme is described. Ligands designed with this multi-target combination in mind are intended as starting points for future development of multi-target drugs. These could be beneficial in treating multi-factorial diseases (as discussed previously), particularly in this case neurodegenerative diseases.[116, 119]

In chapter 4, the method was extended to consider functional effects focusing on compounds, which are simultaneously agonists at A_{2A}R and inhibitors of PDE10A, as a dual-target combination to elevate intracellular cAMP, and provide a proof of concept for the therapeutic benefits that might be exhibited in lung cancer. Triazoloquinazolines were computationally identified as a chemical series that showed agreement in both the ligand- and structure-based predictions of binding to A_{2A}R and PDE10A (which is described in chapter 3). For the purpose of validating this chemical series as dual ligands at these targets, triazoloquinazolines, which are experimentally known PDE10A inhibitors, were docked into the orthosteric site of the A_{2A}R crystal structure. This was performed as a virtual screening step in a structure-based approach that aimed to identify A_{2A}R agonists as part of the dual-target ligand design objective. Subsequently, MD simulations were performed to study the dynamics of a specific amino acid residue in the orthosteric site,[319–321] in order to find a conformational descriptor characterizing A_{2A}R activation. Finally, the anti-proliferative effects of the dual ligands at A_{2A}R and PDE10A in lung carcinoma cell-lines were experimentally measured in order to provide proof of concept for the potential therapeutic benefits that these dual-target ligands might exhibit in lung cancer.

2 Structure-based design of allosteric calpain-1 inhibitors populating a novel bioactivity space

2.1 Introduction

Calpains are proteins that belong to the family of calcium-dependent, non-lysosomal cysteine proteases expressed ubiquitously in mammals and other organisms.[317] Although the physiological roles of calpains are still poorly understood, they have been shown to be involved in many processes such as cell motility, long-term potentiation in neurons and cell fusion in myoblasts.[322] In particular, dimeric calpains have been reported to be involved in cell degeneration processes that characterize numerous disease conditions.[323] The discovery of selective calpain inhibitors however, has been extremely challenging.[316] In this chapter, a structure-based virtual screening protocol, which uses the PEF(S) crystal structure,[18] is reported in order to address this problem. It is shown for the first time that the inhibition of enzyme activity can be attributed to an allosteric mode of action, which may offer improved selectivity and a reduced side-effects profile.

Calpain-1 (μ -calpain) and calpain-2 (m-calpain) are heterodimeric proteases composed of a large subunit with a molecular mass of ~80 kDa, associated with a small subunit of mass ~30 kDa. The small subunit consists of two domains, namely the penta-EF hand calcium binding PEF(S) domain and a glycine rich (GR) domain which are essential for stabilizing calpain-1 and calpain-2.[324] High sequence similarity of 62% is exhibited by the large subunits of calpain-1 and -2 in humans.[324] They consist of four different domains, an N-terminal anchor helix, the active site domain (CysPc), a domain that resembles the C2 membrane binding domains of phosphokinases (known as the C2L domain) and a second penta-EF hand calcium binding domain known as PEF(L). The PEF(L) domain is the determinant of the calcium concentration that is required for protease activation, which differentiates between the two isoforms: calpain-1 requires micromolar concentrations of Ca^{2+} , whereas calpain-2 requires millimolar concentrations for activation.[18]

Calpain-1 is a target dysregulated in many diseases such as neurodegenerative disorders, cardiovascular diseases, ischaemic disorders, arterial sclerosis, leishmaniasis and cancer.[15, 16] In most cases of disease, calpain-1 activity is elevated (and hence its inhibition would be beneficial in treatment). However, it has been recently suggested that the up regulation of

calpain-1 appears to be beneficial in some cases, such as stage II Alzheimer's disease, where its activation could be neuroprotective and beneficial in controlling cellular damage.[325]

Until recently, it has been challenging to design selective calpain-1 inhibitors, and this is attributed to the fact that most compounds that target the active site inhibit a broad spectrum of cysteine proteases, thereby resulting in undesirable side effects.[326] For example, it has been previously reported that calpain-1 inhibitors,[326] which also inhibit the proteasome may induce apoptosis, whereas selective calpain-1 inhibitors do not. Hence, it may be beneficial to design selective calpain-1 inhibitors to avoid off-target related side effects.

Classical allosteric inhibitors of calpain-1, which were originally reported to bind to PEF(S), exhibit a specific type of chemistry - that is α -mercaptoacrylic acid-based, such as compounds PD150606 and PD151746.[327] These inhibitors are potent, cell permeable and selective inhibitors of calpain-1 and calpain-2 exhibiting selectivity towards calpain over other cysteine proteases, with a slight selectivity for calpain-1 over calpain-2. It has been reported however that PD150606 could equally inhibit the active site domain of calpain-1 *without* the presence of PEF(S).[318] This obviously suggests that its mode of action is rather unclear. The reported α -mercaptoacrylic acid based calpain inhibitors (Figure 7A. PD150606 and B. PD151746) and their disulfide analogues (Figure 7C. and D.) were synthesized by Adams *et al*, and shown to bind to PEF(S) from X-ray diffraction analysis (PDB IDs: 1NX3, 4WQ2 and 4WQ3).[18, 328–330] Additional calpain-1 inhibitors that were reported to inhibit the calpain-1 complex, which consists of the PEF(S) and CysPc (active site of calpain-1), are also depicted in Figure 7, including their chemical structures. CHEMBL203568,[17] shown in Figure 7E. is a compound reported to inhibit the calpain-1 complex with an IC_{50} value of 4.9 nM. While CHEMBL204883[17], shown in Figure 7F. is a compound reported to inhibit the calpain-1 complex with an IC_{50} value of 8 nM. However, CHEMBL203568 and CHEMBL204883 have not been confirmed as PEF(S) binders i.e. exhibiting an allosteric mode of inhibition. Their reported confidence score is 7 (in ChEMBL) indicating that these compounds might be binding to any of the subunits involved in the full-length calpain-1 complex. Hence, the use of PEF(S) (calpain-1 small subunit) in structure-based virtual screening may be an appealing approach for the design of allosteric calpain-1 binders with completely different structural architectures from the classical allosteric binders and inhibitors. In addition, this approach is expected to answer the question of whether PEF(S) binding would confer inhibition, given that the shortlisting of candidates is based on the prediction of their binding to PEF(S).

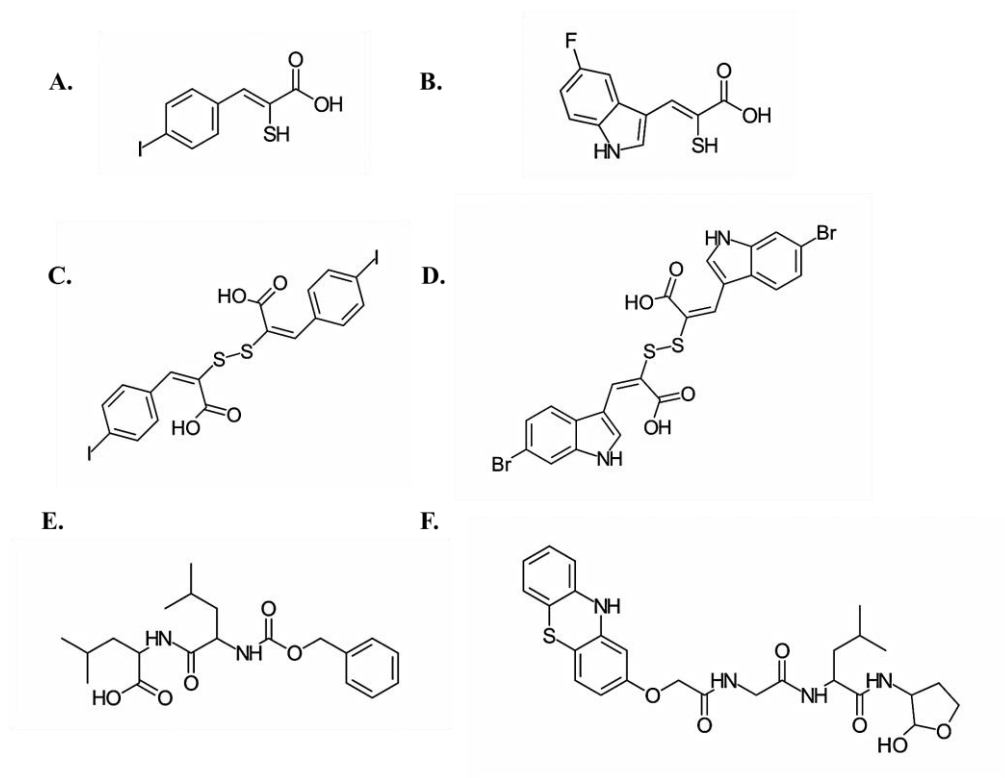


Figure 7. PD150606 **A.** PD151746 **B.** bind to the PEF(S) of calpain, and show modest selectivity for calpain-1 over calpain-2. Adams *et al.*, 2015, synthesized their disulphide analogues **C.** and **D.** respectively, which bind to PEF(S) and were reported with improved potencies in comparison to their monomer compounds. ChEMBL203568 **E.** is a compound reported to inhibit the calpain-1 complex (Uniprot IDs: P04632 and P07384) with an IC_{50} value of 4.9 nM and confidence score of 7. ChEMBL204883 **F.** is a compound reported to inhibit the calpain-1 complex (Uniprot IDs: P04632 and P07384) with an IC_{50} value of 8 nM and confidence score of 7

In this work, the PEF(S) (PDB ID: 4WQ2, calpain-1 small subunit) has been used in a structure-based virtual screening protocol to ascertain whether novel chemical series bind to the allosteric pocket. To validate this approach, purchasable ligands of diverse and novel structural frameworks (very different from those that have been previously investigated) were evaluated *in silico*, using ligand/protein docking against PEF(S), and the compounds were subsequently tested in relevant assays. This pipeline for the structure-based virtual screening protocol is depicted in Figure 8.

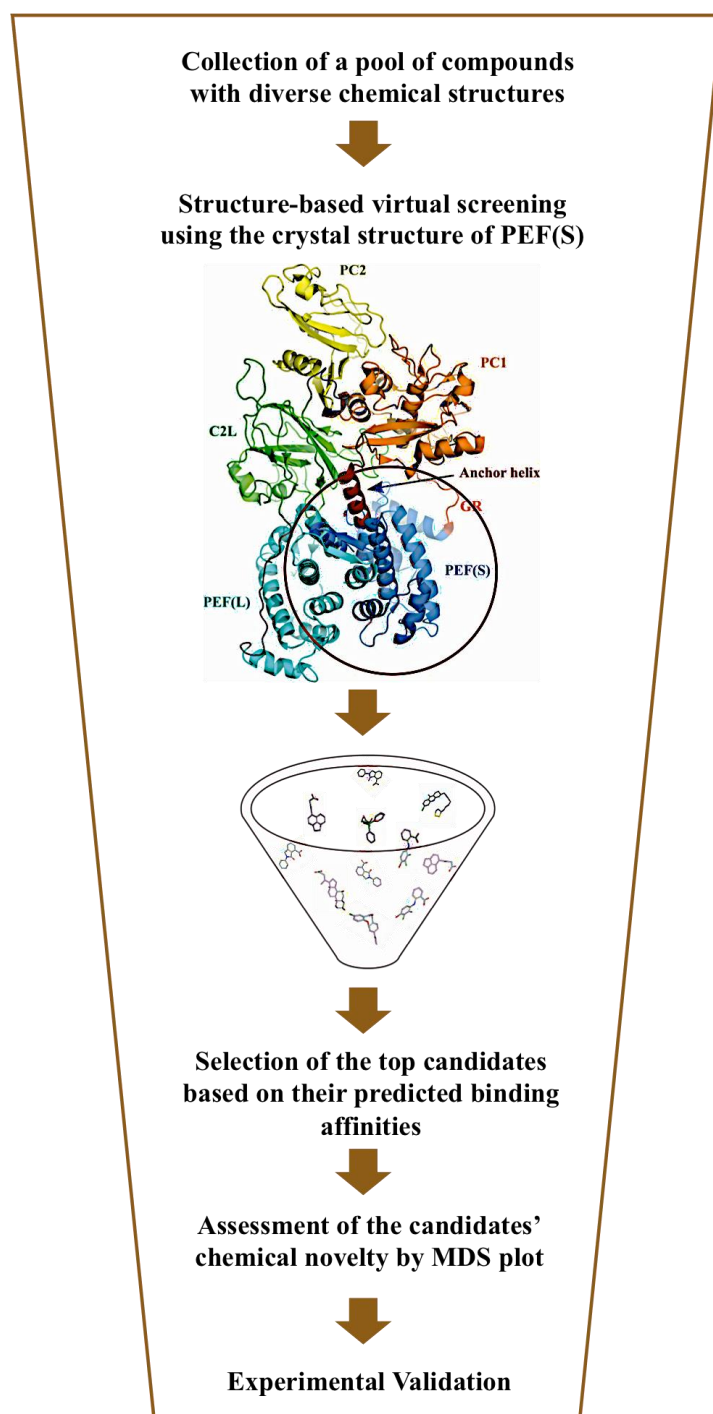


Figure 8. The pipeline of the structure-based virtual screening protocol followed for shortlisting candidates of PEF(S) binders started with the collection of a pool of compounds with diverse chemical structures (very different from those that have been previously investigated), then candidates were shortlisted based on docking into the crystal structure of PEF(S) (PDB: 4WQ2, calpain-1 small subunit). The top ranked candidates were assessed for their chemical novelty in comparison to the classical allosteric binders and inhibitors using an MDS plot. The subset of compounds was then investigated using relevant experimental assays.

As a general approach, it was hypothesized that the functional effect of PEF(S) binders on the active site of calpain-1 may be predicted by carrying out Molecular Dynamics (MD)

simulations[331, 332] on the full-length calpain-1 complex (PEF(S) and CysPc) in both the unbound and the ligand bound states at the PEF(S) domain. If the bound compound increases the average distance between the substrate and the interacting residues in the active site of calpain-1, then it was further postulated that it would inhibit the activity of the enzyme. In contrast, if the compound decreases this distance, it would favor interactions between the enzyme and the substrate, thus facilitating the enzyme reaction. A crystal structure of the full-length calpain-1 complex (PEF(S) and CysPc) is currently unavailable, and hence the problem was approached by using the pipeline shown in Figure 8. Accordingly, the compounds that are predicted to bind to PEF(S) are either expected to inhibit or activate calpain-1.

To explore the small molecule chemical architecture that would most likely alter the geometry of the calpain-1 active site, to inhibit its substrate cleavage allosterically, compounds with a diverse chemical structures were investigated computationally based on their predicted binding affinities towards PEF(S). The candidate PEF(S) binders, which were shortlisted by the structure-based virtual screening protocol depicted in Figure 8 were sulphonamides, N-{3-[3-(2-alkoxyethoxy)-5-(4-substituted-phenyl)-1H-1,2,4-triazol-1-yl]phenyl}-4-methyl-3-(trifluoromethyl)benzamides, and [1,2,4]triazolo[4,3-b]pyridazin-6-yl]pyridines. Experimentally these compounds were also shown to bind to PEF(S) by displacing 2-p-toluidinylnaphthalene-6-sulfonate (TNS, which binds to the allosteric site).[327] In addition, three compounds were able to inhibit the full-length calpain-1 complex (which includes PEF(S) and CysPc) allosterically, but not the active site domain of calpain-1 in the absence of PEF(S). The micro-molar inhibitory activity, *via* a proposed allosteric mechanism is a crucial finding given that these compounds show specificity in their mode of action, which is not the case for the classical allosteric inhibitors such as PD150606. The new inhibitors possess different scaffolds from the classical allosteric inhibitors. These compounds serve as a novel starting point for the expansion of the compound series (including SAR) to improve their potency. In addition, this finding suggests that compounds can inhibit the enzyme activity *via* the PEF(S) domain. An important aspect of this study is that by designing allosteric inhibitors, which do not inhibit the active site domain (that is common to a wide variety of cysteine proteases) these may be effective in treating calpain-1 related diseases without the side effects associated with inhibitors which inhibit the active site domain as well.[316]

2.2 Materials and Methods

2.2.1 Extraction and preparation of purchasable compounds for structure-based virtual screening against PEF(S)

Purchasable compounds (36,503) with diverse chemistry, including sulphonamide-, amide-, pyridine-, urea- and enamine-based compounds, were downloaded from the Aldrich market select database-2016.[333]

The entire set of extracted ligands were prepared for docking with LigPrep 2.5[334] using the default settings and the Epik option, which introduces energy penalties associated with ionization and tautomerization.[335]

2.2.2 Receptor Preparation of PEF(S)

The preparation of the human PEF(S), calpain-1 small subunit (regulatory subunit) of the protein crystal structure (PDB[151] ID: 4WQ2)[18] bound to (Z)-3-(6-bromondol-3-yl)-2-mercaptoacrylic acid was performed for protein-ligand docking with Glide[134]. The structure was prepared using the Protein Preparation Wizard of Maestro 9.3,[334] following the default protocol, which accounts for energy refinement, hydrogen addition, pK_a assignment, and side-chain rotational isomer refinement. Resolved water molecules were discarded and the structure was centered using the co-crystallized ligand as the center of the receptor grid generated for the protein structure. The co-crystal structure of the human calpain PEF(S) protein crystal structure (PDB ID: 4WQ2) bound to (Z)-3-(6-bromondol-3-yl)-2-mercaptoacrylic acid was selected as the target structure.

2.2.3 Cut-off generation for compound selection from docking model

In an attempt to validate the docking model, a set of known active and inactive compounds were docked against the PEF(S) protein crystal structure to ensure that it enriched actives. 32 compounds manually extracted from ChEMBL[171] with IC₅₀ values ≤ 1 μ M (protein complex of the calpain-1, catalytic and small regulatory subunits: P07384, P04632 with confidence scores of 6 or 7) were docked against the PEF(S) model. In addition, 20 inactive compounds of the PEF(S) calpain-1 small regulatory subunit (Uniprot ID: P04632), extracted from PubChem (using the SQL query in Appendix D),[336] were docked.

A good separation was obtained for the medians of the docking score distributions for actives versus inactives for the docking model indicating that the actives are enriched. Figure 9 shows

the separation of the medians for the PEF(S) docking model, the medians are -7.48 (actives) vs -4.60 (inactives). In addition, the indole and the phenyl α -mercaptoacrylic acid-based inhibitors and their disulfide analogues (33 compounds), which were synthesized by Adams *et al.*[18] and shown by X-ray crystallography to bind to PEF(S), were docked against the PEF(S) docking model. The median of the docking score distribution obtained is -7.25 in comparison to the median of inactives, which is -4.60 (Figure 10). This further indicates that the model enriches these set of actives. A Mann-Whitney test, which included statistical analysis on the active and inactive docking score distributions, was performed with R[337] using the script provided by Kalash *et al.*[338] The differences in the medians was significant with p values less than 0.05.

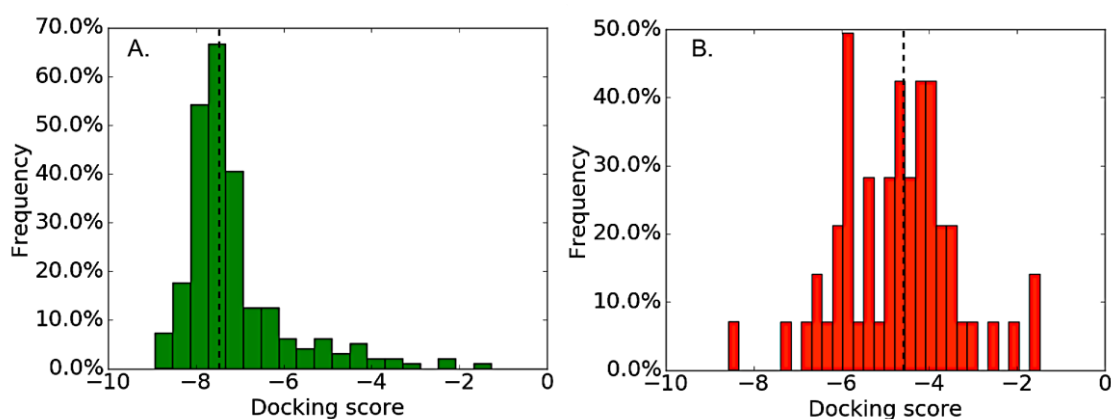


Figure 9. A good separation was obtained for the medians (dashed lines) of the docking score distribution for active versus inactive compounds for the calpain-1 docking model **A.** -7.48 (active compounds) vs. **B.** -4.60 (inactive compounds). Statistical analysis was performed using a Mann-Whitney test on the docking score distributions of each target. The difference in medians was significant (p value < 0.05), indicating that the ChEMBL actives are enriched.

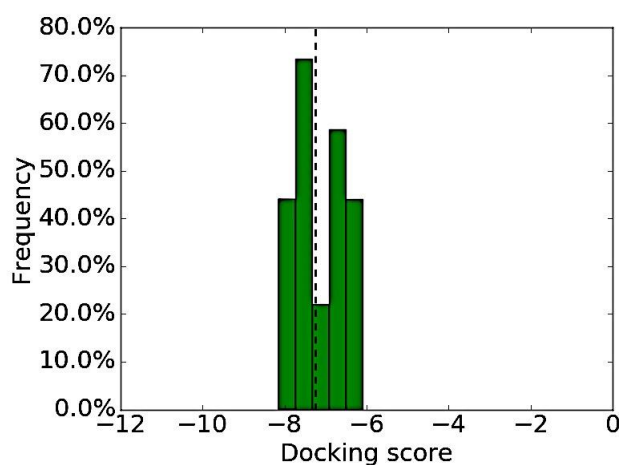


Figure 10. The median (dashed line) of the docking score distribution obtained for the α -mercaptoacrylic acid based inhibitors and their disulfide analogs, which were synthesized by

Adams *et al.*,[18] was -7.25. The difference between this median and the median of PubChem inactives was significant (p value < 0.05), indicating that the actives of the Adams library are enriched.

The Matthews correlation coefficient (MCC), which takes into account true and false positives and negatives, was computed using a Python script[338] for all the docking scores of the ChEMBL actives and PubChem inactives for the model. A search was performed for a docking score threshold that gave the highest MCC for the docking model. This has been done in order to shortlist purchasable candidates as PEF(S) binders, which displayed docking scores that are more negative than the docking score threshold with the highest MCC for the PEF(S) docking model (which was -6.35).

2.2.4 Docking

The purchasable compounds, prepared according to the protocol described in section 2.2.1, were docked into the PEF(S) protein crystal structure (PDB ID: 4WQ2). The Glide docking parameters included extra precision (XP) and the flexible ligand sampling option, which were deduced from docking experiments using known actives and inactives against the protein model. The highest ranked compounds with respect to their predicted affinity towards PEF(S) were selected for binding assessment and calpain-1 activity evaluation i.e. those which displayed docking scores that are more negative than the score with the highest MCC for the PEF(S) docking model, which was -6.35. The compounds that did not exhibit potential PAINS[339] liabilities (with regard to the recent analysis of the use of this approach (Tropsha))[313, 340] upon virtual screening with the FAFDrug3 ADME-Tox filtering tool, were selected for experimental validation. The shortlisted compounds exhibited diverse structures (depicted in Table 1), with five sulphonamides **1-5**, two substituted N-{3-[3-(2-alkoxyethoxy)-5-(4-substituted-phenyl)-1H-1,2,4-triazol-1-yl]phenyl}-4-methyl-3-(trifluoromethyl)benzamides **6-7**, and three substituted [1,2,4]triazolo[4,3-b]pyridazin-6-yl]pyridines **8-10**.

2.2.5 Multi-Dimensional Scaling (MDS) analysis of the shortlisted compounds 1-10

This step of the analysis aimed to plot the chemical space of ChEMBL compounds which are active against the full-length calpain-1 complex (with IC₅₀ values ≤ 1 μM (protein complex of calpain-1, catalytic and small regulatory subunits: P07384, P04632 and confidence scores of 6 or 7)), and the Adams library,[18] which were all validated against the PEF(S) docking model. In addition, the shortlisted candidates from the structure-based design protocol were included

in the same plot. This analysis enables an assessment of the novelty of the chemical space coverage of the shortlisted compounds **1-10**. For this purpose, the SMILES of all compounds involved were standardized using the ChemAxon Command-Line Standardizer, where the following options were selected: “Remove Fragment” (keep largest), “Neutralize”, “RemoveExplicitH”, “Clean2D”, “Mesomerize” and “Tautomerize”.[341]

Subsequently, Morgan fingerprints (radius 2, 1024 bits) were generated for all the compounds using KNIME 2.11.3.[342] The workflow generated Morgan fingerprints in the following sequence: It (a) read chemical data from an SDF file, (b) generated RDKit molecules from a molecule string representation (SDF), (c) generated hashed bit-based fingerprints for an input RDKit Mol column (d) converted RDKit molecules into string based molecule representations (SDF or SMILES) (e) excluded columns from the input table (f) renamed columns (g) and saved data table into a CSV file.

A 2D-similarity matrix based on Euclidean distance of generated Morgan fingerprints was computed using the `dist()` function in R.[337] Then, a multidimensional scaling metric of the similarity matrix was computed by embedding it into two dimensions ($k=2$). Then, for each data set in the plot (corresponding to the Adams library, ChEMBL compounds, and the shortlisted compounds), 90% confidence ellipses were computed using the ellipse package. Finally, based on the generated Morgan fingerprints, a 2D MDS plot with 90% ellipse-like confidence regions was obtained using the R `ggplot2` package,[343] with the x-axis and the y-axis labeled Dimension 1 and Dimension 2 respectively. These are two relative and unit-less dimensions that recapitulate the pairwise similarity of the all points observed in the distribution of Euclidean distances in a lower dimensional space.

2.2.6 Experimental validation of the virtual screening protocol

The experimental validation of the structure-based design of PEF(S) binders was performed by Joel Cresser-Brown (JCB) and Connor Morgan (CM) (in the group of Professor Allemann) at Cardiff University, and their experimental protocols are described below.

2.2.6.1 Expression and purification of PEF(S)

The codon optimized gene encoding human PEF(S) was purchased from Epoch Biolabs (Texas, USA) in a pET21d vector. Human PEF(S) was produced in *E. coli* BL21-CodonPlus (DE3)-RP (Agilent Technologies), and purified using the same procedure previously described for PEF(S) (work done by JCB).[18]

2.2.6.2 Evaluating calpain-1 activity of the shortlisted candidates of PEF(S) binders 1-10

This assay uses a fluorogenic peptide from the calpain-1 substrate α -spectrin, containing a FAM-DABCYL FRET pair ($\text{H}_2\text{N-K(FAM)-EVYGMMK(DABCYL)-OH}$). Cleavage by calpain-1 occurs between the Tyr-Gly residues and results in enhanced fluorescence as the quenching effect is relieved. The assays using purified porcine calpain-1 (CalBiochem, 25 nM) were performed in a buffer containing 1 μM calpain-1 substrate, 10 mM HEPES, 10 mM DTT, 0.5 mM EDTA, bovine serum albumin (0.1%) pH 6.8. The assay was carried out using a fluorescent plate reader (BMG Optistar) with a final assay volume of 100 μl at a temperature of 37°C, using an excitation band pass filter centered at 490 nm and emission detected at 520 nm. The compounds were added to the assay mixture before the reaction was initiated by the addition of CaCl_2 (5 mM). None of the compounds had significant fluorescence at this wavelength. The compounds were dissolved in DMSO at 40 mM and diluted into assay buffer to give range of concentrations from 5 nM to 200 μM . In each assay run, the effect of DMSO alone over the concentration used was also measured. Although there was no effect of DMSO at lower concentrations, in some assay runs, DMSO at 0.005%-0.5% produced some inhibitory effect. This DMSO effect (which was only relevant for compounds with poor inhibitory ability) was subtracted before constructing the inhibition curves.[18] The IC_{50} values were obtained by fitting the data with non-linear regression using the SigmaPlot software,[344] and the reported results are the mean \pm standard deviation of three independent experiments (work done by JCB and CM).

2.2.6.3 TNS displacement for compounds 1-5

10 μM PEF(S) in 20 mM Tris base, 1.1 mM CaCl_2 , 1 mM EDTA and pH 7.4 was incubated with 46.7 μM 2-p-toluidinylnaphthalene-6-sulfonate (TNS, 1mM stock in 40% ethanol) in a Greiner CELLSTAR 96 well black flat bottom plate for 5 minutes at 25 °C.

Compounds **1-10** were stored as 40 mM stock solutions in DMSO, then diluted from 500 μM to 500 nM by serial dilution over 10 wells with an Integra Viaflow 96 multichannel pipette in triplicate.

The plates were then incubated in a FLUOstar Omega plate reader at 25 °C for 5 minutes then were analyzed using an excitation wavelength of 355 nm and an emission wavelength of 450 nm with 10 flashes per well with orbital averaging.

The baseline (just TNS) was subtracted from the fluorescence (B). B was then subtracted from the B_{\max} (with no inhibitor) to invert the data. The data was plotted in excel as bar diagrams for the mean \pm standard deviation of three independent experiments ($n=3$) in triplicate. Whereas for compounds **6-10**, their fluorescence interfered with TNS, (refer to Figure 11. for the blanks obtained for compounds **1-10**) for this reason their results were omitted (work done by JCB).

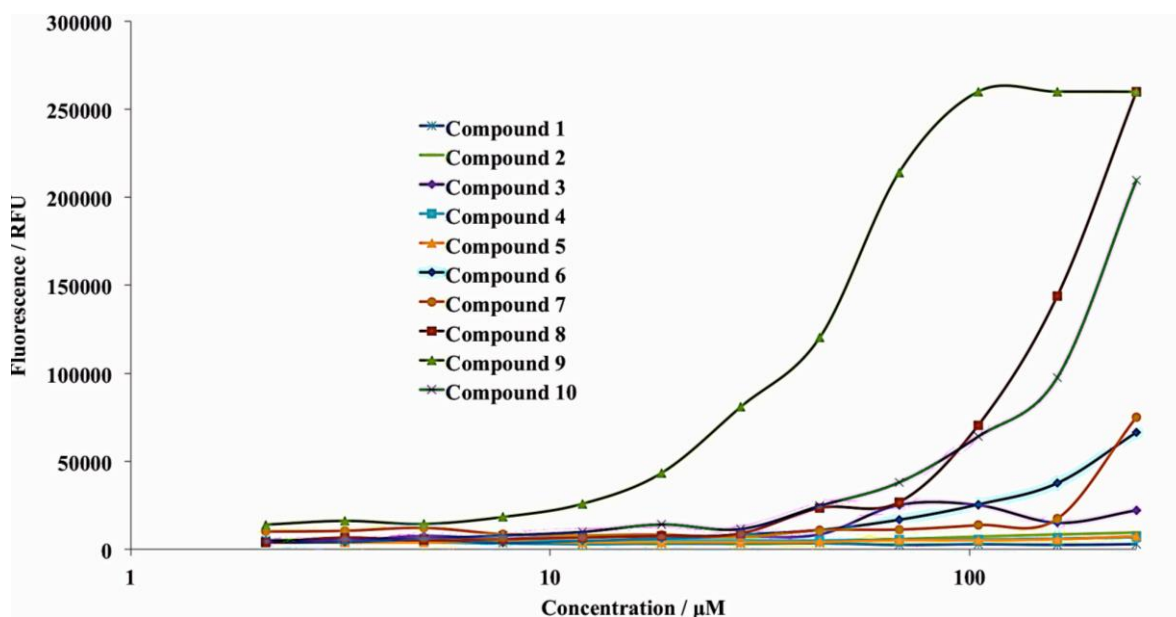


Figure 11. Blanks for compounds **1-10** measured at the TNS excitation and emission wavelengths, 355 and 450 nm respectively. It appeared that compounds **6-10** interfered with the fluorescence of TNS so they were not subject to the TNS displacement method to assess their binding to the PEF(S) domain.

Compounds and reagents Compound **1** was purchased from Tocris, and compounds **2-10** were purchased from Ambinter, and used without further purification. PD150606 was purchased from Sigma Aldrich and used without further purification.

2.2.6.4 Cloning of CysPC gene

The pET28a-GB1-MAAKLVFF plasmid, a kind gift from Dr Cornelius Krasel of the Institut für Pharmakologie (Marburg, Germany), was used to create a Golden Gate acceptor plasmid by standard PCR, overlap extension PCR, endonuclease digestion and T4 DNA ligase reactions. When subjected to Golden Gate digestion/ligation with BsaI and T4 DNA ligase with an appropriate complimentary PCR product, the resulting plasmid has a section of DNA encoding RF under a constitutive promotor removed and the PCR product is incorporated in such a fashion that the translation product contains the protein of interest with an N-terminal

hexahistidine tag GB1 fusion tag that can be removed by the action of tobacco etch virus protease (TEV) (work done by JCB).

Digestion/ligation proceeds as expected despite the presence of an additional BsaI site within the sequence encoding RFP. The CysPC domain of calpain-1 was inserted into this acceptor by PCR with the following primers (work done by JCB):

Fwd: CGACTAGTGGTCTCCAGTCCATGGGTCGCCATGAGAA

Rev: CGACTAGTGGTCTCCATCAGTCCGGGGTCAGGTTACA for ligation using the golden gate protocol into the pET28a-GB1-MAAKLVFF plasmid.[345]

Plasmid Sequencing

T7 Fwd

TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCAT
CATCATCATCACACTTACAAATTAATCCTTAATGGTAAAACATTGAAAGGCGAAAC
AACTACTGAAGCTGTTGATGCTGCTACTGCAGAAAAAGTCTTCAAACAATACGCT
AACGACAACGGTGTTGACGGTGAATGGACTTACGACGATGCGACTAAGACCTTTA
CAGTTACTGAACATATGGAAAACCTGTATTTTCAGTCCGAGACCTTTACGGCTAGC
TCAGCCCTAGGTATTATGCTAGCTACTAGAGAAAGAGGAGAAAACTAGTATGGTT
AGCAAAGGCGAGGAGCTGATTAAGGAGAATATGCACATGAAACTGTACATGGAAG
GCACCGTGAACAACCACCACTTCAAGTGCACCAGCGAGGGTGAAGGCAAACCGT
ATGAAGGCACCCAGACCATGCGTATCAAAGTGGTTGAGGGTGGCCCGCTGCCGTT
CGCGTTTGATATTCTGGCGACCAGCTTCATGTACGGTAGCCGTACCTTTATCAACCA
CACCCAGGGCATTCCGGATTTCTTTAAACAGAGCTTCCCGGAAGGTTTTACCTGGG
AGCGTGTGACCACCTACGAAGACGGTGGCGTTCTGACCGCGACCCAGGACACCA
GCCTGCAAGATGGCTGCCTGATCTATAACGTGAAGATTCTGTGGTGTAACTTTCCG
AGCAACGGCCCCGGTGATGCAGAAGAAAACCCTGGGTTGGGAGGCGAACACCGA
AATGCTGTATCCGGCGGATGGTGGCCTGGAGGGCCGTAGCGACATGGCGCTGAAG
CTGGTTGGTGGCGGTCACCTGATCTGCAACTTCAAAACCACC

T7 Rev

CGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTTATC
AGGAGACCGCTAGTTCAGTTTGTGACCCAGCTTGCTCGGCAGATCGCAATAACGC
GCAACCGCCACTTCGTGTTGCTCAACGTAGGTCTCTTTATCCGCTTCCTTAATACGC

TCCAGACGGTGATCAACATAGTACACACCCGGCATTTCAGGTTCTTCGCCGGTTT
 CTTGCTACGATAGGTGGTTTTGAAGTTGCAGATCAGGTGACCGCCACCAACCAGC
 TTCAGCGCCATGTCGCTACGGCCCTCCAGGCCACCATCCGCCGGATACAGCATTTC
 GGTGTTTCGCCTCCCAACCCAGGGTTTTCTTCTGCATCACCGGGCCGTTGCTCGGA
 AAGTTAACACCACGAATCTTCACGTTATAGATCAGGCAGCCATCTTGCAGGCTGGT
 GTCCTGGGTTCGCGGTCAGAACGCCACCGTCTTCGTAGGTGGTCACACGCTCCCAG
 GTAAAACCTTCCGGGAAGCTCTGTTTAAAGAAATCCGGAATGCCCTGGGTGTGGT
 TGATAAAGGTACGGCTACCGTACATGAAGCTGGTCGCCAGAATATCAAACGCGAA
 CGGCAGCGGGCCACCCTCAACCACTTTGATACGCATGGTCTGGGTGCCTTCATACG
 GTTTGCCTTCACCCTCGCTGGTGCACCTGAAGTGGTGGTTGTTACGGTGCCTTCC
 ATGTACAGTTTCATGTGCATATTCTCCTTAATCAGCTCCTCGCCTTTGCTAACCATAC

2.2.6.5 Expression and purification of CysPC

BL21-CodonPlus (DE3) RP cells containing the human calpain-1 CysPC gene were grown at 37 °C in kanamycin selective LB media until $OD_{600} = 0.6$ then induced with 1 mM IPTG. The protein was expressed overnight at 20 °C and cells harvested by centrifugation in a Sorvall RC6 Plus centrifuge (Thermo Fisher Scientific, Inc, MA, USA) using an SLA-3000 rotor at 6080 RCF for 20 minutes at 4 °C. The cells were re-suspended in 20 mM HEPES, 100 mM NaCl, 0.5 mM TCEP pH 7.6 (buffer A) and lysed by sonication for 5 mins (pulsed 5 s on, 10 s off). The lysate was clarified by centrifugation at 4 °C for 40 minutes at 30310 RCF in a Sorvall RC6 Plus centrifuge. The supernatant was passed through a 0.2 µm syringe filter and applied to a Ni-NTA column. The bound protein was washed with 15 CV buffer A and eluted with 10 CV buffer A containing 250 mM imidazole, which was further dialyzed in buffer A overnight in a 10 kDa membrane containing 1 mL aliquot of TEV protease. The cleavage product was then passed back through a Ni-NTA column to remove the 6xHis-GB1 solubility tag and TEV protease, with the flow through containing active CysPC as confirmed by SDS-PAGE, mass spectrometry and calpain-1 activity assay (work done by JCB).

2.2.6.6 Expression and purification of TEV protease

BL21 (DE3) cells containing the TEV gene codon optimised for *E. coli* expression were obtained from Prof. Nigel Richards (Cardiff University).

The cells containing the TEV protease gene were grown at 37 °C in ampicillin selective LB media until $OD_{600} = 0.6$ then induced with 1 mM IPTG. The protein was expressed overnight

at 20 °C and cells harvested by centrifugation in a Sorvall RC6 Plus centrifuge (Thermo Fisher Scientific, Inc, MA, USA) using an SLA-3000 rotor at 6080 RCF for 20 minutes at 4 °C. The cells were re-suspended in 20 mM HEPES, 100 mM NaCl, 0.5 mM TCEP pH 7.6 (buffer A) and lysed by sonication for 5 mins (pulsed 5 s on, 10 s off). The lysate was clarified by centrifugation at 4 °C for 40 minutes at 30310 RCF in a Sorvall RC6 Plus centrifuge. The supernatant was passed through a 0.2 µm syringe filter and applied to a Ni-NTA column. The bound protein was washed with 15 CV buffer A and eluted with 10 CV buffer A containing 250 mM imidazole. The eluent was mixed with 20% v/v glycerol (20 mL final volume), and stored at -80 °C in 1 mL aliquots (work done by JCB).

2.3 Results and Discussion

2.3.1 Structure-based virtual screening of purchasable ligands against PEF(S)

36,503 commercial compounds consisting of diverse chemical structures including sulphonamide-, amide-, pyridine-, urea-, and enamine-based compounds, were docked using Glide into the pre-prepared (see methods for details) protein crystal structure of human PEF(S) (PDB[151] ID: 4WQ2).[18] From the docking scores, the distribution for actives versus inactives was obtained. The active molecules displayed a more favorable distribution of scores, which allowed differentiation of actives and inactives (see methods for details).

Candidate PEF(S) binders from the purchasable database were shortlisted on the basis of a cut-off with the highest Mathews Correlation coefficient. The cut-off obtained was -6.35, according to which compounds with more negative binding score were predicted to bind. The selected candidates were further screened against PAINs[339] using the FAFDrug3 ADME-Tox Filtering Tool.[313] Those compounds that didn't exhibit any potential PAINs liability were considered for evaluation of calpain-1 activity. As a result, five sulphonamides **1-5**, two substituted N-{3-[3-(2-alkoxyethoxy)-5-(4-substituted-phenyl)-1H-1,2,4-triazol-1-yl]phenyl}-4-methyl-3-(trifluoromethyl)benzamides **6-7**, and three substituted [1,2,4]triazolo[4,3-b]pyridazin-6-yl]pyridines **8-10**, which were the top ranked compounds according to the virtual screening criteria were shortlisted as candidates for PEF(S) binding.

2.3.2 MDS plot shows that shortlisted PEF(S) binders occupy a novel region in chemical space

An MDS plot of the chemical space was generated consisting of the ChEMBL compounds inhibiting the full-length calpain-1 complex (PEF(S) and CysPc) with IC₅₀ values ≤ 1 µM, and

the Adams library (a library of α -mercaptoacrylic acid-based calpain-1 inhibitors and their disulfide analogues),[18] which were all validated against the PEF(S) docking model (see methods for details) (Figure 12). In addition, the shortlisted candidates **1-10** from the structure-based design protocol were also included in the plot, which enabled the assessment of the novelty of chemical space coverage of the shortlisted PEF(S) binders, in comparison to the classical calpain-1 inhibitors and PEF(S) binders. Figure 12, which is a two dimensional MDS plot based on Morgan fingerprints of radius 2 with 90% ellipse-like confidence regions, shows that the shortlisted compounds **1-10** exhibited new structures in comparison to the previously reported compounds by occupying a novel region in the chemical space of calpain-1 actives.

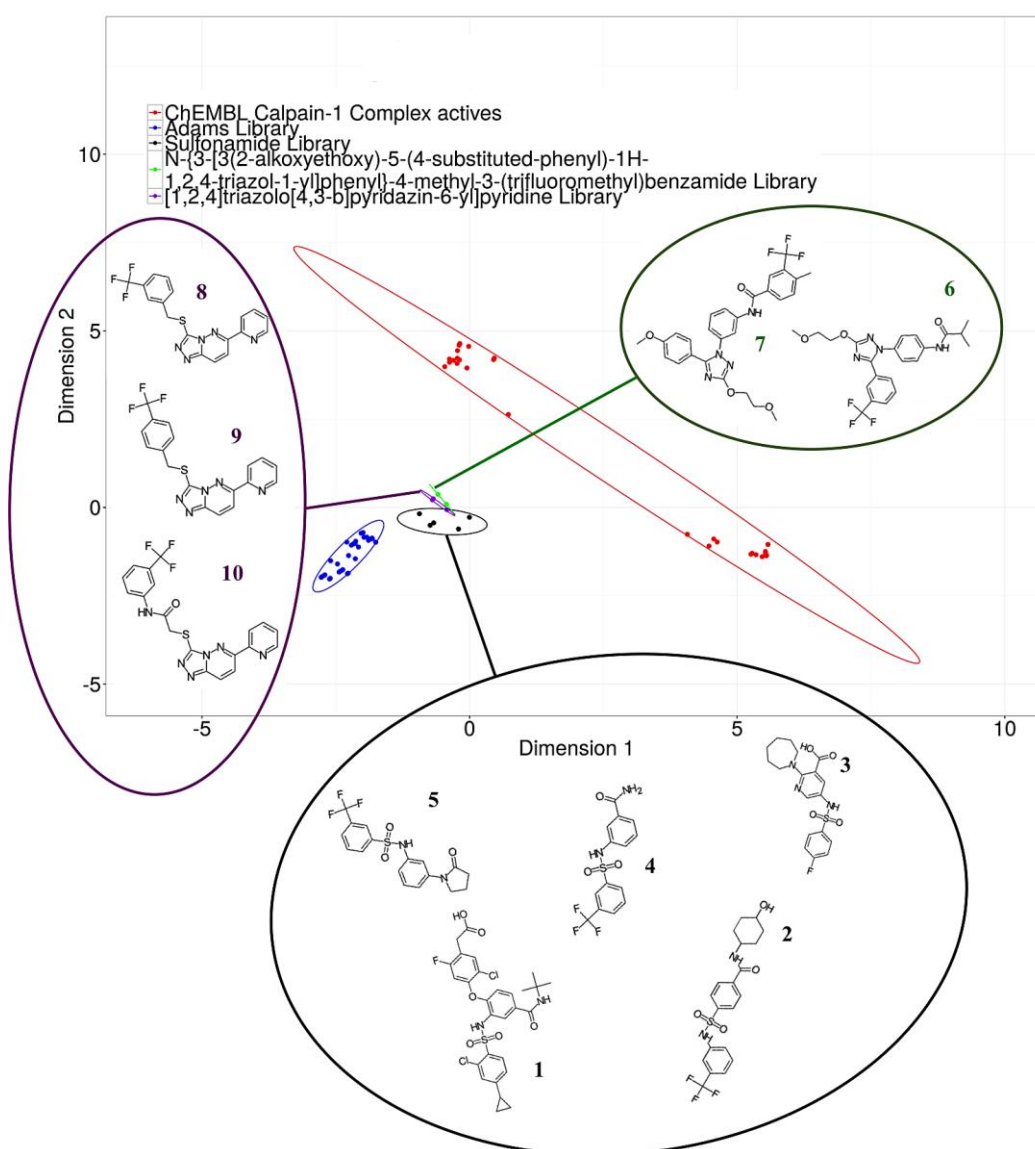


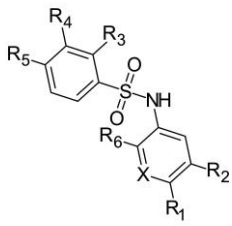
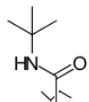

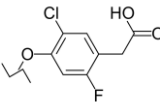
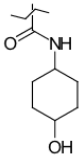
Figure 12. A two dimensional MDS plot based on Morgan fingerprints of radius 2, with 90% ellipse-like confidence regions for 32 ChEMBL compounds, 33 compounds synthesized by Adams *et al.*, [18] and the 10 shortlisted candidates from the structure-based design protocol. These were shortlisted from a purchasable database based on their high predicted binding




affinity towards the PEF(S) domain of calpain *via* docking with Glide. The compounds identified belong to a new chemical space in comparison to the previously reported compounds that bind to the PEF(S) (Adams library) and the compounds which inhibit the full-length calpain-1 complex. Compounds **1**, **9**, and **10** were then experimentally tested as novel allosteric inhibitors of calpain-1.


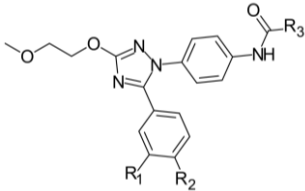
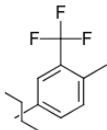
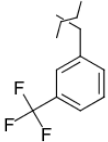
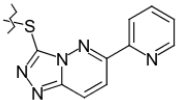
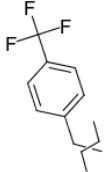
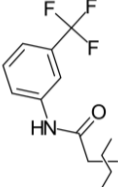
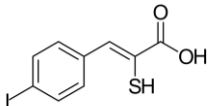
2.3.3 FRET based Inhibition Assay

FRET based inhibition experiments were carried by JCB and CM. A fluorogenic assay of the full-length calpain-1 complex (which includes PEF(S) and CysPc) was used to determine the activity of compounds **1-10**. Amongst the identified compounds, **1**, **9**, and **10** inhibited the full-length calpain-1 complex with IC₅₀ values of 7.5 (±1.1), 20.5 (±1.9), and 29.7 (±5.2) μM, respectively (Table 1). The same experimental protocol was performed to measure the activity of compounds against the active site domain of calpain-1, without the presence of PEF(S), to investigate a possible allosteric mode of action. None of the compounds showed any inhibition in the absence of PEF(S), except for compound **1**, which weakly inhibited the active site domain with an IC₅₀ value >100 μM. In contrast, compound **3** exhibited higher inhibitory activity against the active site domain of calpain-1 with an IC₅₀ value of 41.1 (±15.4) μM as compared to the full-length calpain-1 complex, which showed an activity of >100 μM, suggesting that in the presence of PEF (S) it preferentially binds to PEF(S). This could explain the reduction in the inhibitory activity of compound **3** since it is most likely unable to alter the geometry of the active site while it binds allosterically. The IC₅₀ values were also measured for the classical α-mercaptoacrylic acid based calpain inhibitor PD150606, and these were 19.3 (±1.6) μM for the full-length calpain-1 complex, and 17.8 (±2.4) μM with the active site domain without the presence of PEF(S). Hence, in contrast to compounds **1**, **3**, **9**, and **10**, PD150606 exhibited an unspecific mode of action by equally inhibiting *via* both binding sites (active and allosteric sites). It is worth stating here that compound **1** is an asthma drug, Vidupiprant or AMG 853,[346] which exhibited higher potency than PD150606 in inhibiting the activity of calpain-1. The dose response curve of compound **1** (IC₅₀ = 7.5±1.1 μM) is displayed in Figure 13. Interestingly, identifying Vidupiprant as an allosteric inhibitor of calpain-1 is in agreement with previous reports showing a direct link between calpain inhibition and anti-inflammatory properties, where it was shown that non-steroidal anti-inflammatory drugs (NSAIDs) inhibit calpain, and that calpain inhibition reduces allergic inflammation.[347, 348] Potentially, this finding could highlight the importance of considering calpain inhibitors for the development of new anti-asthma therapies.

Table 1. IC₅₀ values for compounds **1-10**, and PD150606 determined by the FRET based inhibition assay, with the full-length calpain-1 complex and the active site domain of calpain-1, reported as mean +/- standard deviations from three independent experiments (NR = no response)

Compound	Common Scaffold	Substituents	IC ₅₀ full-length calpain-1 complex (μM)	IC ₅₀ active site domain calpain-1 (μM)
1		<p>R₁ and R₄ = H</p> <p>R₂ = </p> <p>R₃ = Cl</p> <p>R₅ = </p> <p>R₆ = </p> <p>X = C</p>	7.5 ± 1.1	>100
2		<p>R₁, R₃, R₄, and R₆ = H</p> <p>R₂ = CF₃</p> <p>R₅ = </p> <p>X = C</p>	NR	NR

3		$R_1 = $  $R_2 = \text{CO}_2\text{H}$ $R_3, R_4, \text{ and } R_6 = \text{H}$ $R_5 = \text{F}$ $X = \text{N}$	>100	41.1±15.4
4		$R_1, R_3, R_5 \text{ and } R_6$ $=\text{H}$ $R_2 = $  $R_4 = \text{CF}_3$ $X = \text{C}$	NR	NR
5		$R_1, R_3, R_5 \text{ and } R_6$ $=\text{H}$ $R_2 = $  $R_4 = \text{CF}_3$ $X = \text{C}$	NR	NR
6		$R_1 = \text{CF}_3$ $R_2 = \text{H}$	NR	NR

		$R_3 =$ 		
7		$R_1 = H$ $R_2 = OCH_3$ $R_3 =$ 	NR	NR
8			NR	NR
9			20.5 ± 1.9	NR
10			29.7 ± 5.2	NR
PD150606			19.3 ± 1.6	17.8 ± 2.4

The allosteric inhibitory activity exhibited by compounds **1**, **9**, and **10** confirms the design approach, which shortlisted PEF(S) binders. The three active compounds include scaffolds that

are distinct from the classical allosteric inhibitors. The IC_{50} values obtained could be improved by efficient choice of substituents using standard medicinal chemistry approaches. In particular compound **1**, which is a sulphonamide, exhibited specificity in its allosteric inhibition of calpain-1, and was more potent ($7.5\ \mu\text{M}$) than PD150606 ($19.3\ \mu\text{M}$) that additionally inhibits the active site domain, in the absence of PEF(S), with a similar IC_{50} value.

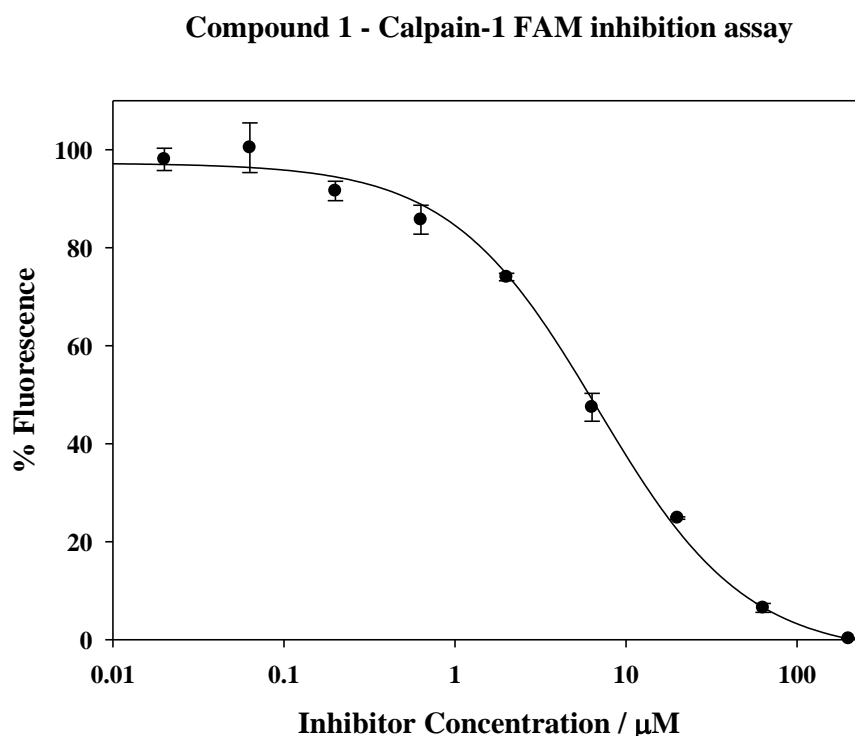


Figure 13. Dose-response curve for the inhibition of full-length calpain-1 complex by compound **1**, $IC_{50} = 7.5 \pm 1.1\ \mu\text{M}$. The IC_{50} value is reported as mean \pm standard deviation of three independent experiments (plot generated by JCB).

2.3.4 TNS Displacement Assay

TNS, which is a sensitive fluorophore that binds to PEF(S) was used to probe protein dynamics and conformational change. It fluoresces in the bound state i.e. in a hydrophobic environment, whereas when another compound displaces it, its fluorescence gets quenched. This fluorophore was used previously to assess the binding of PD150606 to PEF(S), a compound which has already been shown to bind to PEF(S) by X-ray crystallography (PDB ID: 1NX3).[327, 328] In this work, PD150606 (as a control), compound **1**, the most potent allosteric inhibitor, compound **3**, a weak allosteric inhibitor, and compounds **2**, **4** and **5** which did not exhibit any inhibitory activity, have been tested (by JCB) for PEF(S) binding by the TNS displacement method. The results for compounds **6-10** were unreliable as it became apparent that they

fluoresced under the assay conditions, therefore for this reason these results were omitted. As shown in Figure 14, all tested compounds, except for compound **5**, quenched the fluorescence of TNS, exhibiting a similar trend in their quenching effect to that of the known PEF(S) binder, PD150606, hence confirming their binding to PEF(S). Therefore, it appears that compounds **2-4** do indeed bind to PEF(S) in a similar fashion to compound **1**, but they either weakly inhibited or failed to allosterically alter the geometry of the active site.

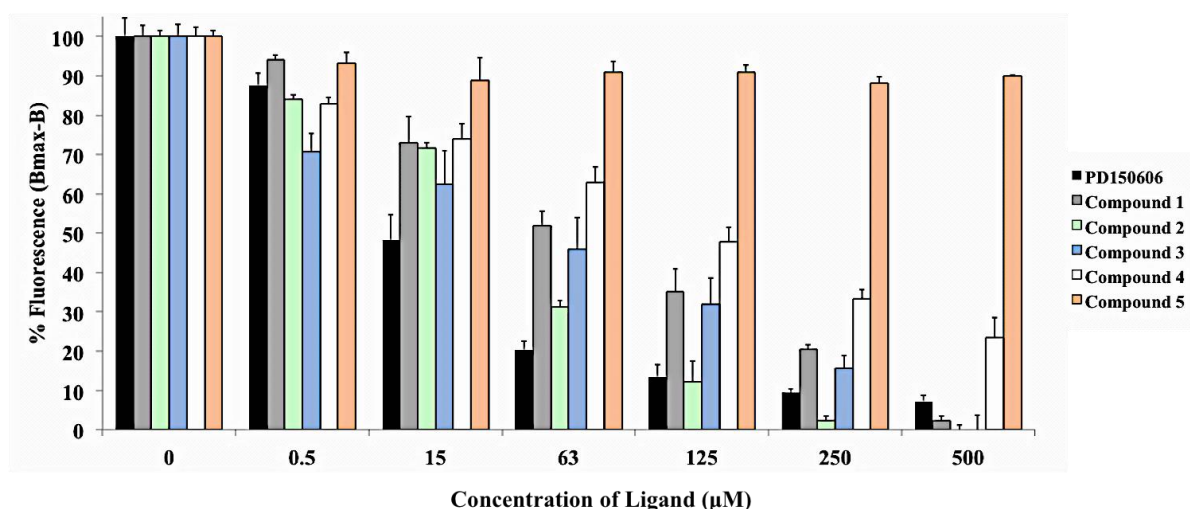


Figure 14. PD150606, which was shown by X-ray crystallography to bind to PEF(S) (PDB ID: 1NX3), compound **1** (the most potent allosteric inhibitor), **3**, which was the least potent among the identified allosteric inhibitors, **2** and **4**, which did not exhibit any inhibitory activity, all quenched the fluorescence of TNS. All compounds showed a similar effect to that of PD150606 confirming their binding to PEF(S), except for compound **5**, which neither exhibited any inhibitory activity nor displaced TNS.

2.3.5 Analysis of molecular docking studies of representative calpain-1 inhibitors **1** and **10**, and compounds **2-5**

Docking studies predicted molecular interactions of the sulphonamide **1** and the [1,2,4]triazolo[4,3-b]pyridazin-6-yl]pyridine **10** with the PEF(S) protein crystal structure (PDB ID: 4WQ2). Figure 15A. shows the 2-chloro-4-cyclopropylsulfonamido phenyl ring of compound **1** is π -stacked with His₁₃₁. The carbonyl of its carboxylic acid moiety H-bonds with same residue. The carbonyl of its amide moiety H-bonds with Trp₁₆₈, and the phenyl ring attached to the tert-butylcarbamoyl moiety is π -stacked with the same residue. Figure 15B. shows a π -stacking interaction between the pyridine ring of compound **10** and Trp₁₆₈, and H-bonding of the nitrogen in that ring with the same residue. The hydrophobic interactions predicted for compounds **1** and **10** with Trp₁₆₈ are also seen in the co-crystallized ligand/protein crystal structure (PDB ID: 4WQ3).[18] In addition, a more favorable binding affinity towards PEF(S) was predicted for compound **1** ($IC_{50} = 7.5 \pm 1.1 \mu M$) as compared to compound **10**,

($IC_{50} = 29.7 \pm 5.2 \mu M$), which could be the reason for the higher inhibitory activity exhibited by compound **1**. In order to further explore the activities of compounds **1-10**, MD simulations on the full-length calpain-1 complex, which includes the PEF(S) domain would be beneficial when (if) the crystal structure is available. As previously intimated, the inhibitory activity of each compound is predicted to correlate with the average distance between the substrate and the interacting residues in the active site of calpain-1.

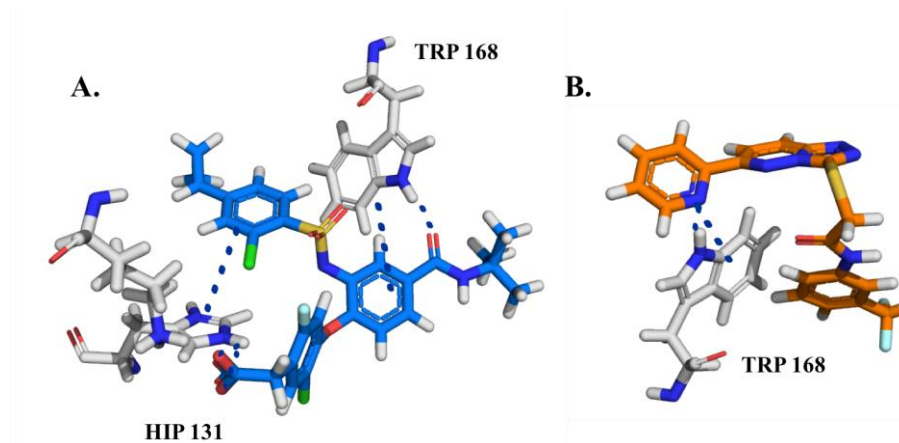


Figure 15. Docking studies predicted molecular interactions of compounds **1** and **10** with the human PEF(S) of calpain-1 small subunit (regulatory subunit) protein crystal structure (PDB ID: 4WQ2). **A.** The 2-chloro-4-cyclopropylsulfonamido phenyl ring of compound **1** shows π -stacking with His₁₃₁ and the carbonyl of its carboxylic acid moiety H-bonds with the same residue. The carbonyl of its amide moiety H-bonds with Trp₁₆₈ and the phenyl ring attached to the tert-butylcarbamoyl moiety is π -stacked with the same residue **B.** The pyridine ring of compound **10** is π -stacked with Trp₁₆₈ and the nitrogen of that ring H-bonds with same residue. The hydrophobic interactions of **1** and **10** with Trp₁₆₈ are also seen in the co-crystallised ligand/protein crystal structure (PDB ID: 4WQ3).[18]

Docking studies predicted molecular interactions of the sulphonamides **2-5** with the PEF(S) protein crystal structure (PDB ID: 4WQ2), which suggest that the hydrophobic interactions with Trp₁₆₈ are essential for PEF(S) binding. Figure 16A. has the phenyl ring of compound **2**, which is attached to the sulfonyl moiety, π -stacked with Trp₁₆₈, and the same residue H-bonds with the carbonyl of the amide group. The hydroxyl of its cyclohexyl ring H-bonds with Gln₁₀₀, and its sulfonyl group H-bonds with His₁₃₁. Figure 16B. shows the predicted molecular interactions between compound **3** and the PEF(S) crystal structure. These are H-bonding between its carboxyl moiety and the Lys₁₇₂ and Trp₁₆₈ residues, H-bonding between its sulfonyl group and His₁₃₁, and π -stacking with the Trp₁₆₈ *via* its pyridine ring. Figure 16C. demonstrates the predicted molecular interactions for compound **4**. The amino group of its amide moiety H-bonds with the Glu₉₇ and Trp₁₆₈ residues, the phenyl ring of its benzamide group π -stacks with Trp₁₆₈, and the carbonyl of its amide moiety H-bonds with Lys₁₇₂. Figure 16D. shows the H-

bonding interactions predicted for the carbonyl of the oxopyrrolidin-1-yl moiety in compound **5** to the Lys₁₇₂ and Trp₁₆₈ residues, the H-bonding interactions of its sulfonyl moiety with His₁₃₁, and π -stacking of the same residue with the aromatic ring of its trifluorophenyl moiety. Interestingly, the hydrophobic interactions predicted for compounds **1-4** with Trp₁₆₈ are also seen in the co-crystallized ligand/protein crystal structure (PDB ID: 4WQ3).[18] However, compound **5** was predicted to exhibit hydrophilic interactions with Trp₁₆₈, which could explain why it didn't displace TNS, suggesting that the hydrophobic interactions with this residue are essential for PEF(S) binding.

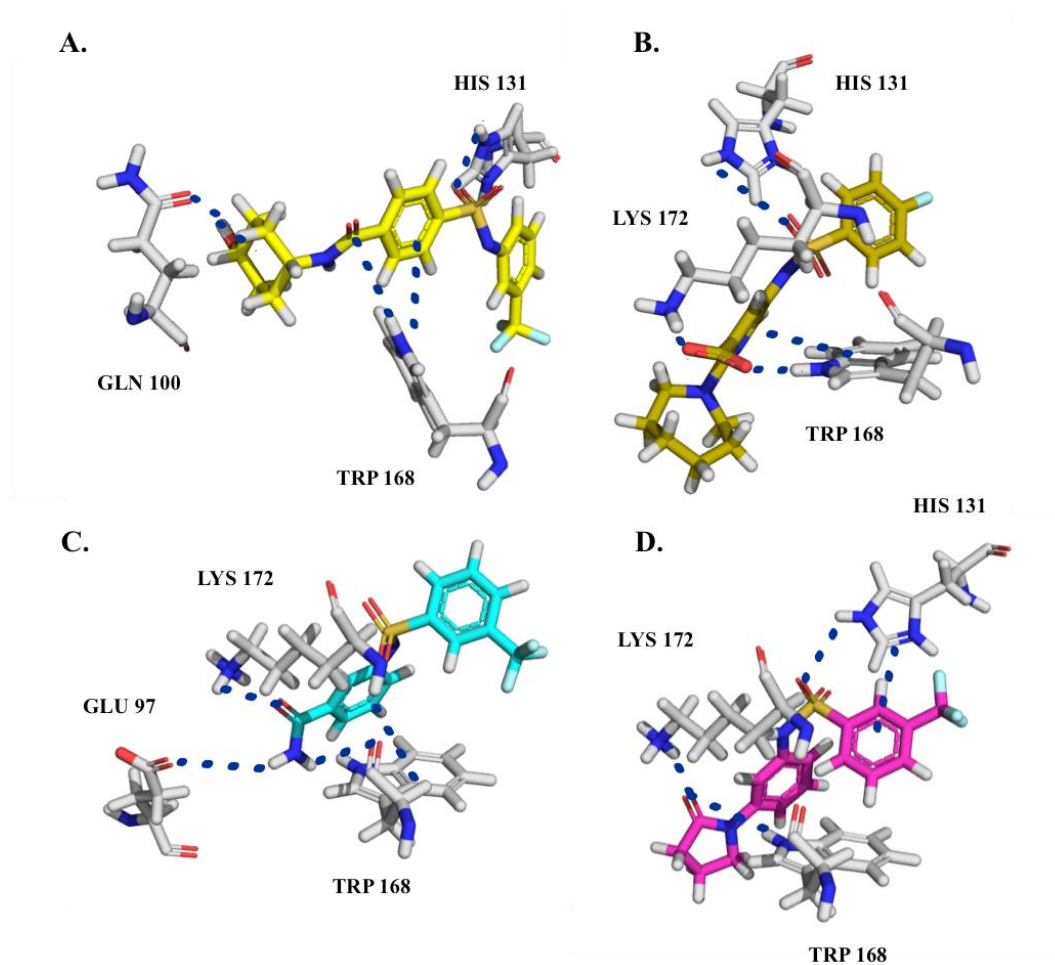


Figure 16. Docking studies predicted molecular interactions of the sulphonamides **2-5** with the PEF(S) protein crystal structure (PDB ID: 4WQ2), which suggest that hydrophobic interactions with Trp₁₆₈ are essential for PEF(S) binding **A.** The phenyl ring of compound **2**, which is attached to the sulfonyl moiety is π -stacked with Trp₁₆₈, and the same residue H-bonds with the carbonyl of its amide group, the hydroxyl of its cyclohexyl ring H-bonds with Gln₁₀₀, and its sulfonyl group H-bonds with His₁₃₁ **B.** H-bonding interactions are predicted to occur between the carboxyl moiety of compound **3** and the Lys₁₇₂ and Trp₁₆₈ residues, H-bonding between its sulfonyl group and His₁₃₁, and π -stacking with the Trp₁₆₈ *via* its pyridine ring **C.** The amino group of the amide moiety for compound **4** H-bonds with Glu₉₇ and Trp₁₆₈, and the phenyl ring of its benzamide group π -stacks with Trp₁₆₈, and the carbonyl of its amide moiety

H-bonds with Lys₁₇₂ **D.** H-bonding interactions are predicted to occur between the carbonyl of the oxopyrrolidin-1-yl moiety for compound **5** and the Lys₁₇₂ and Trp₁₆₈ residues, H-bonding between its sulfonyl moiety and His₁₃₁, and π -stacking of the same residue with the aromatic ring of its trifluorophenyl moiety.

2.3.6 Computational assessment of CNS permeability for representative calpain-1 inhibitors **1** and **10**

Given that calpain-1 may serve as a therapeutic target for neurodegenerative disorders, a computational assessment of the CNS permeability for compounds **1** and **10** was performed with FAFDrug3[313] to see whether these compounds could be considered as good starting points to target these diseases. Their physicochemical properties were calculated and CNS diagrams were obtained and are presented in Figure 17A. and B. Compound **1** did not pass the CNS filter, which takes into consideration the assessment of its ability to pass the blood brain barrier. Hence, it is predicted not to exhibit the desired permeability,[304] since the values of all the descriptors for compound **1** (logP, HBD, HBA, MW-molecular weight, tPSA (blue line)) fall outside the CNS filter area (light blue). As for compound **10**, it is predicted to exhibit medium permeability since all the descriptors, except for the HBA, have passed the CNS filter. Hence, compound **10** might serve as a good starting point for analogue development.

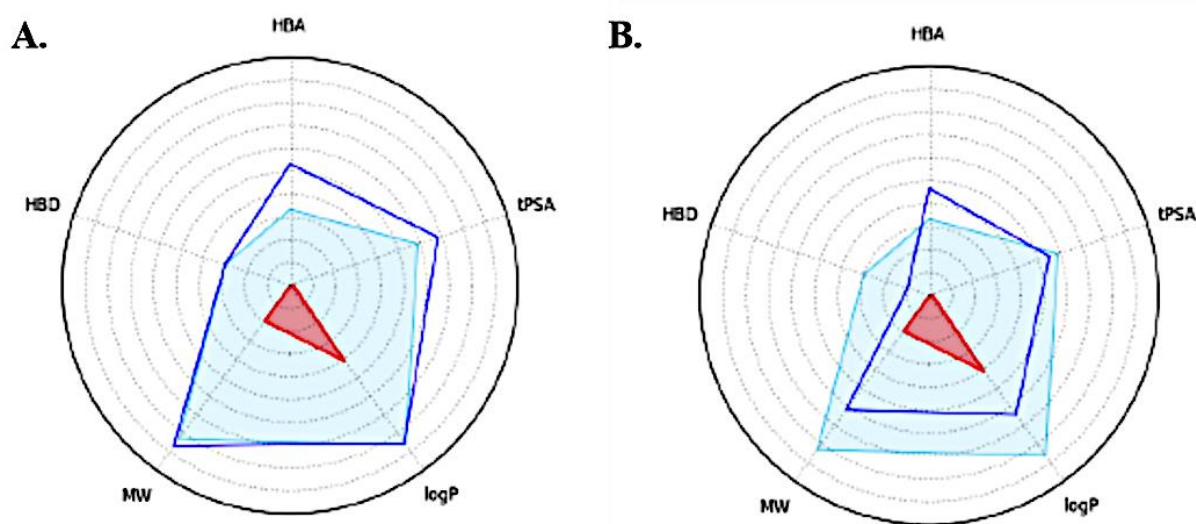


Figure 17. Compounds **1** and **10** are represented in diagrams **A** and **B**. **1** did not pass the CNS filter which takes into consideration the assessment of its ability to pass the blood brain barrier, values of descriptors for **1** (the descriptors logP, HBD, HBA, MW, tPSA (blue line)) fall outside the CNS filter area (light blue). For **10**, it is predicted to exhibit medium permeability since only the HBA value lies outside the CNS filter area.

2.4 Conclusions

In this work, the structure-based design method devised has been successfully validated, which has led to the discovery of chemically novel allosteric inhibitors of calpain-1, for the first time demonstrating an allosteric mode of action. Compounds **1**, **9**, and **10** inhibited the full length calpain-1 complex (which includes PEF(S) and CysPc) with IC₅₀ values of 7.5 (\pm 1.1), 20.5 (\pm 1.9), and 29.7 (\pm 5.2) μ M respectively. Compounds **9** and **10** did not inhibit the active site domain of calpain-1 in the absence of PEF(S), and compound **1** inhibited the active site domain weakly with an IC₅₀ value >100 μ M. In contrast, compound **3** exhibited higher inhibitory activity against the active site domain of calpain-1 with an IC₅₀ value of 41.1 (\pm 15.4) μ M as compared to the full-length calpain-1 complex, which was >100 μ M, suggesting that it preferentially binds to PEF(S). In addition, IC₅₀ values were measured for PD150606, giving 19.3 (\pm 1.6) μ M with the full-length calpain-1 complex and 17.8 (\pm 2.4) μ M with the active site domain (without the presence of PEF(S)). In comparison to the classical α -mercaptoacrylic acid based calpain inhibitor, PD150606, compounds **1**, **9**, and **10**, exhibited specificity in their allosteric mode of action, since they didn't inhibit the active site domain in the absence of PEF(S).

Furthermore, PD150606, compound **1**, the most potent allosteric inhibitor, compound **3**, a weak allosteric inhibitor, and compounds **2**, **4** and **5** (which did not exhibit any inhibitory activity) have been tested for PEF(S) binding by the TNS displacement method. Compounds (**1-4**) quenched the fluorescence of TNS, exhibiting a similar trend in their quenching effect to that of the known PEF(S) binder, PD150606.

The micro-molar IC₅₀ values obtained for compounds **1**, **9**, and **10** suggest that they may be good starting points for optimization, having novel scaffolds. Allosteric inhibitors discovered by this approach could exhibit more selectivity towards calpain-1 since they are unlikely to inhibit the active site domain, which is similar for a wide variety of cysteine proteases. This could translate to more effective treatments with less side effects for calpain-1 related diseases.[316]

3 Structure- and ligand-based design of multi-target ligands at A₁R, A_{2A}R and PDE10A- key proteins in neurodegenerative diseases

3.1 Introduction

Compounds designed to display polypharmacology may have utility in treating complex diseases, where activity at multiple targets is required to produce a clinical effect. In particular, suitable compounds may be useful in treating neurodegenerative diseases by promoting neuronal survival in a synergistic manner *via* their multi-target activity at the adenosine A₁ and A_{2A} receptors (A₁R and A_{2A}R) and phosphodiesterase 10A (PDE10A), which modulate intracellular cAMP levels.[98, 349–351] In this chapter, a computational method for the design of synthetically feasible ligands that bind to A₁ and A_{2A} receptors and inhibit phosphodiesterase 10A (PDE10A) is described. The method involves a retrosynthetic approach employing *in silico* target prediction and docking,[133, 134, 159, 352] which may be generally applicable to multi-target compound design at several target classes.

Neurodegeneration involves the progressive loss of the structure and function of neurons, which is common in Parkinson's, Huntington's disease and schizophrenia.[353] Recently, there has been substantial interest in the search for alternative non-dopamine (non-DA) based approaches for the treatment of neurodegenerative diseases, as the classical DA-based approaches have long been associated with many undesirable side effects such as dyskinesia, hallucinations, and on/off effects.[354] Given that the adenosine neuromodulation system (*via* the adenosine A₁ and A_{2A} receptors) has been identified as a key target for the management of neurodegenerative diseases, *via* the targeted modulation of cAMP levels, this qualifies its targeting as a potential non-DA based treatment approach.[349, 355] Indeed, modulation of cAMP levels has proven to have benefits in neuronal survival in an adenosine receptor-dependent manner.[356] In addition, recent findings suggest that phosphodiesterase 10A (PDE10A) also plays a role in neurodegenerative diseases with similar etiology such as Parkinson's, Huntington's disease, and schizophrenia.[117–119] Inhibition of PDE10A resulting in the maintenance of elevated intracellular cAMP concentrations has been suggested to be effective in the treatment of these diseases. Thus multi-target ligands that bind to different

adenosine receptors subtypes (A_1 and A_{2A} receptors) while simultaneously inhibiting PDE10A might be synergistic in modulating cAMP levels, and of therapeutic potential [98, 115, 116] In this work, a computational strategy is offered for the design of synthetically feasible ligands that bind to A_1R and $A_{2A}R$, and inhibit PDE10A - a novel multi-target combination of G protein-coupled receptors (GPCRs) and an enzyme, which has not been previously exploited. The designed ligands with this multi-target combination are intended as starting points for future development of multi-target drugs treating neurodegenerative diseases. It should be noted here that the current study considers only affinity of ligands to the above receptors, which are also experimentally validated as outlined below. However, for therapeutically relevant purposes, functional effects and optimization of selectivity towards A_1R , $A_{2A}R$ and PDE10A also need to be considered, - which will be the area of future study.

The workflow of the current study is shown in Figure 18. Starting with a focused chemical space consisting of known actives against A_1R , $A_{2A}R$ and PDE10A, new synthetically feasible compounds were identified *via* RECAP (Retrosynthetic Combinatorial Analysis Procedure),[159, 352] which fragments molecules at pre-defined bonds and recombines them in a combinatorial manner. These were then evaluated *in silico*, using target prediction and ligand/protein docking. Compounds with favorable assessments in both steps were carried forward for substructural analysis. This analysis identified compound series with the highest frequency of prediction as multi-target ligands against the desired set of targets, which has the practical advantage of synthetic accessibility *via* a common synthetic route.

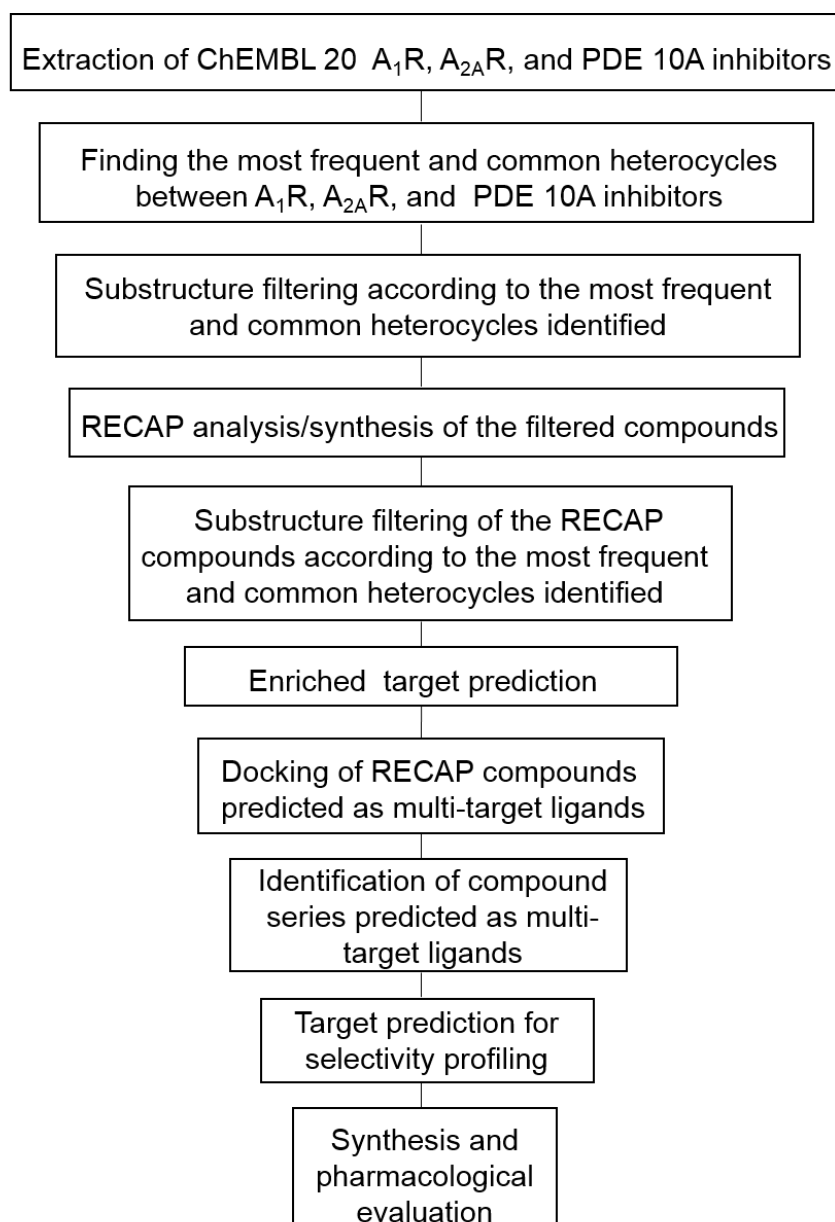


Figure 18. The computational strategy for rational design of A₁R/A_{2A}R-PDE10A multi-target ligands started with a focused chemical space consisting of known actives of A₁R, A_{2A}R and PDE10A. Then formed new synthetically feasible compounds, which were subjected to target prediction and docking for synthesis and pharmacological evaluation.

A series of 2-aminopyridine-3-carbonitriles were selected for prospective validation of the pipeline, a series that was synthetically accessible *via* a one pot synthetic scheme i.e. providing products with the desired properties: cost-effective, synthetically efficient and available in a timely fashion.[357, 358]

Subsequently the synthesized compounds were experimentally tested and confirmed as A₁R/A_{2A}R-PDE10A multi-target ligands. Selectivity against other subtypes of both protein families (A_{2B}R, A₃R, PDE7A, PDE7B, and PDE9A) was assessed due to the high degree of

conservation of residues between the adenosine receptor subtypes in their orthosteric sites and PDEs in their active sites.[351, 359–363] Additionally the pharmacological profile of the compound series was confirmed, and structure activity relationships (SAR) were also deduced. Hence, in this work a successful computational strategy is reported, which allowed the discovery of the first A₁R/A_{2A}R-PDE10A multi-target ligands.

The novel A₁R/A_{2A}R-PDE10A ligands are thought to display an additive effect in modulating the A₁R, A_{2A}R, and PDE10A targets simultaneously similar to that of combination compounds of Adenosine receptors and PDEs, reported by Rickles *et al*, which were synergistic in modulating cAMP levels.[98]

3.2 Materials and Methods

3.2.1 Method for the selection of reference molecules for the design of multi-target ligands

Using an SQL script,[338] human A₁R (2,860), A_{2A}R (3,566) ligands and PDE10A inhibitors (843) were extracted from the ChEMBL 20 database with K_i and IC₅₀ values less than or equal to 1 μ M respectively, and confidence scores of 8 or 9.[171] Following extraction, the most frequent and common heterocycles between A₁, A_{2A} receptor ligands and PDE10A inhibitors were found by performing substructure analysis on each structure using the “Chemistry->Analyze scaffolds” function in DataWarrior 4.2.2.[364] Analysis of A₁R, A_{2A}R ligands and PDE10A inhibitors identified common and frequent heterocycles (pyridine, 1H-pyrazole, pyrimidine and 9H-purine for A₁R and A_{2A}R), and these were extracted from each set using RDKit, 9.1.[365] It should be noted that compounds containing 9H-purine were also extracted from the original set even though this substructure is characteristic of A₁R and A_{2A}R only. These are structurally similar to the common and frequent heterocycles identified (pyridine, 1H-pyrazole, and pyrimidine). Figure 19 shows the most frequent heterocycles for the A₁R, A_{2A}R ligands, and PDE10A inhibitors and their relative frequencies in each set. It was found that they are furan, pyridine, xanthine, 1H-pyrazole, pyrimidine, piperazine, and 9H-purine. All of these heterocycles ranked among the top 30 for A₁R, A_{2A}R ligands and PDE10A inhibitors. This indicated their suitability for the design of multi-target ligands at these protein targets, given the overlap in chemical (heterocyclic) space. In the case where no percentage is displayed for a particular target, this means that the heterocycle does not appear among the top 30 for the set of compounds involved.

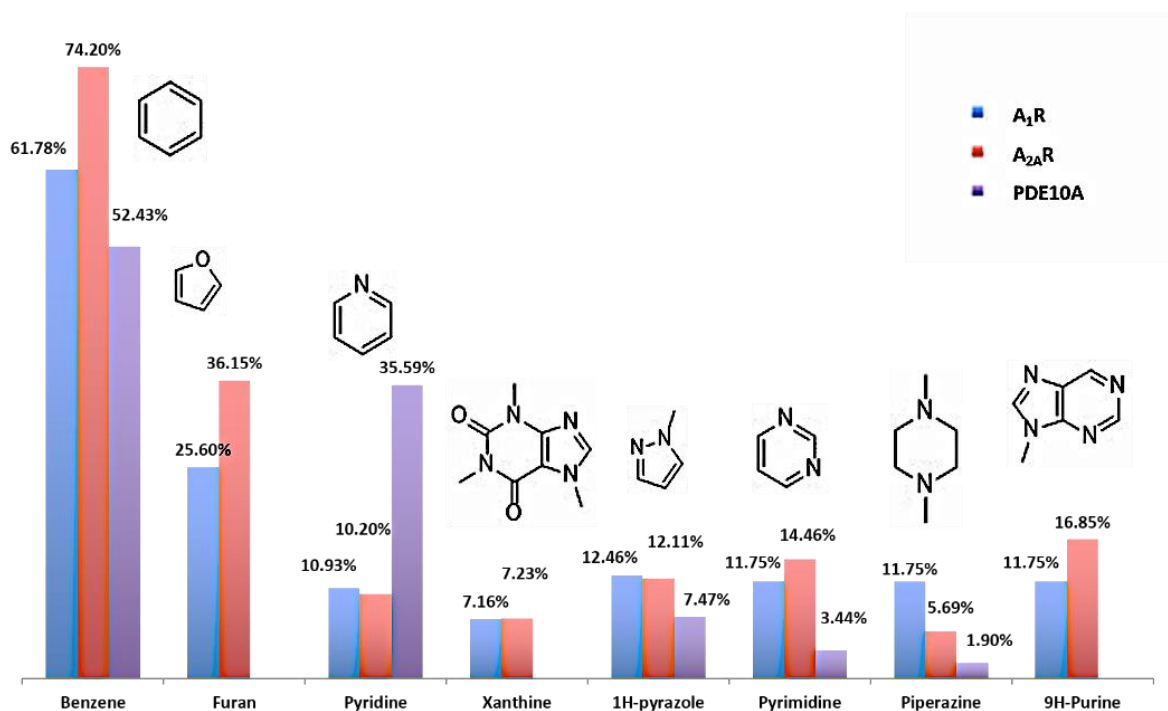


Figure 19. Percentage occurrence of the most frequent substructures in the A₁R, A_{2A}R, and PDE10A ChEMBL compounds (K_i and $IC_{50} \leq 1 \mu M$) reveals the following substructures for A₁R, A_{2A}R, and PDE10A inhibitors: benzene, furan, pyridine, xanthine, 1H-pyrazole, pyrimidine, piperazine, and 9H-purine. In cases where no percentage is displayed for a particular target, this means that the substructure does not appear among the top 30 for the set of compounds involved.

3.2.2 Designing new multi-target ligands

A₁R (2,104), A_{2A}R (2,489) and PDE10A inhibitors (679) consisting of the common and frequent heterocycles, were subjected to RECAP analysis/synthesis in MOE.[352] The RECAP function electronically fragments and recombines molecules based on chemical knowledge of 11 chemical bond types derived from common chemical reactions.[159] As a result, 458,839 novel RECAP-derived compounds were found. Finally, the designed RECAP library was filtered using RDKit library according to the common and frequent heterocycles identified, which narrowed the list down to 22,233 compounds.

3.2.3 Target prediction

The SMILES of the designed RECAP library were standardized using the ChemAxon Command-Line Standardizer where the following options were selected: “Remove Fragment” (keep largest), “Neutralize”, “RemoveExplicitH”, “Clean2D”, “Mesomerize” and “Tautomerize”.[341] The standardized canonical SMILES were exported to CSV files, and subjected to enriched target prediction using PIDGIN 1.0 implementing the method developed

by Liggi *et al.*[133, 366] The target prediction for the designed RECAP library was performed using a recall probability threshold of 0.01 (which is a value consistent with greater confidence in the more positive predictions).

Enrichment calculations for the predicted targets of the designed RECAP library were performed as a complementary approach in order to assess the likelihood of the active compounds against the targets of interest. In this procedure, the frequency of predicting A₁R, A_{2A}R and PDE10A targets for the designed RECAP library was compared with a background distribution of a diverse library covering a large chemical space and was assessed by two parameters: the estimation score and the average ratio. The cutoff selected for considering a target as sufficiently enriched required an estimation score less than or equal to 0.01.[366] The statistical relevance of the prediction was assessed *via* a Chi-squared test with Yates correction in Scipy,[367] using the contingency table of the RECAP library and a background of randomly sampled PubChem compounds (Figure 20).

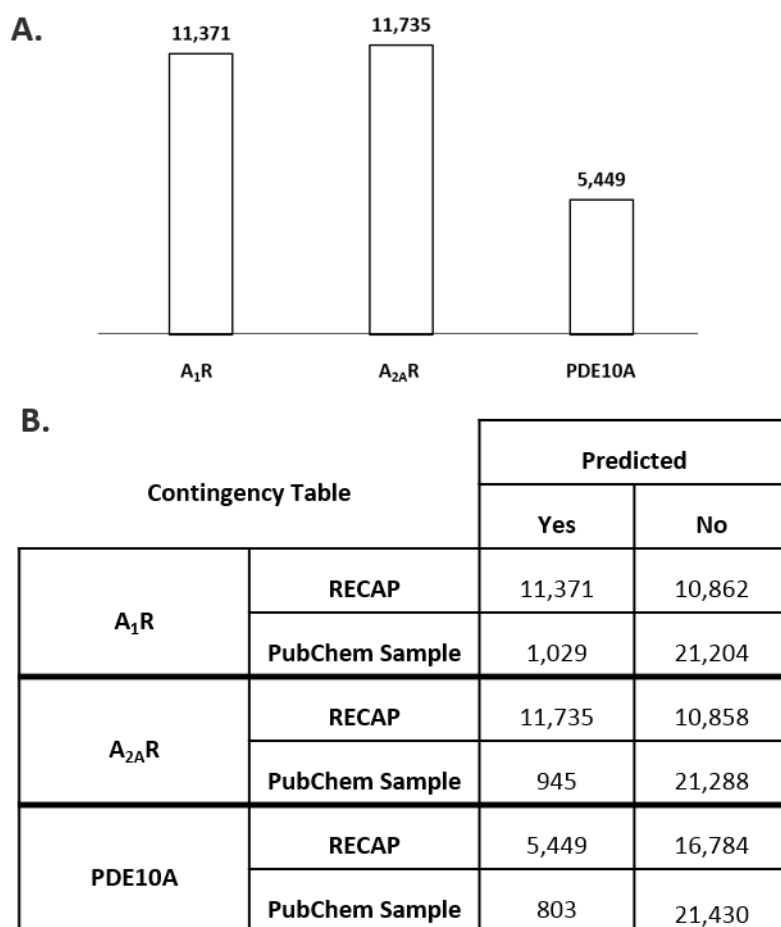


Figure 20. A. A₁R, A_{2A}R and PDE10A were predicted as enriched targets with an estimation score equal to 0 (enriched) and average ratios less than 0.1 (enriched) for the focused RECAP

library, where the percentage of RECAP compounds that were predicted as actives against the A₁R, A_{2A}R and PDE10A targets are: 51.1%, 52.8%, and 24.5%, with Chi-squared p values < 0.005 and Chi-squared statistics of 11958.8, 12842.1, and 4015.7, respectively **B**. The results of the contingency table for the Chi-squared calculation are passed to a Scipy[367] i.e. the script for the calculation for A₁R is `scipy.stats.chi2_contingency ([[1029, 21204], [11371, 10862]])`.

3.2.4 Receptor preparation

The human A_{2A}R protein crystal structure (PDB ID: 4EIY) bound to the antagonist ZM 241385 and the PDE10A crystal structure (PDB ID: 4DDL) complexed with an inhibitor,[368, 369] were selected for docking with Glide[134]. Protein structures were prepared using the Protein Preparation Wizard of Maestro 9.3,[334] following the default protocol, which accounts for energy refinement, hydrogen addition, pKa assignment, and side-chain rotational isomer refinement. Resolved water molecules were discarded, and the structure was centered using the co-crystallized ligand as the center of the receptor grid generated for each protein structure. The co-crystal structures of A_{2A}R with 4-{2-[(7-amino-2-furan-2-yl)[1,2,4]triazolo[1,5-a][1,3,5]triazin-5-yl)amino]ethyl}phenol (PDB ID: 4EIY), and PDE10A with 2-{1-[5-(6,7-dimethoxycinnolin-4-yl)-3-methylpyridin-2-yl]piperidin-4-yl}propan-2-ol (PDB ID: 4DDL), were selected as target structures.

The A₁R homology model was provided by Dr. Hugo Gutiérrez-de-Terán, and it was constructed according to the method reported by Yaziji *et al*,[283, 370, 371] where the protein sequence of the human A₁R (accession number P30542) was aligned with the A_{2A}R template of PDB ID: 4EIY.

3.2.5 Ligand Preparation

The entire set of 2,563 ligands was prepared for docking with LigPrep 2.5[372] using the default settings and the Epik option, which introduces energy penalties associated with ionization and tautomerization.[335]

3.2.6 Cut-off generation for compound selection from docking models

In an attempt to validate the A_{2A}R, A₁R, and PDE10A docking models, a set of known actives and inactives were docked against each target to ensure that they enriched actives. 81 A_{2A}R receptor ligands reported in the literature were docked against the A_{2A}R model.[373, 374] For consistency 81 ChEMBL actives were also extracted manually (for each of the A₁R and PDE10A proteins whose K_i and IC₅₀ values are less than 10 μM), and these were docked against

their respective target class. In addition, PubChem inactives (200 compounds) of each target class were docked.

A good separation was obtained for the medians of the docking score distributions for actives versus inactives confirming that the actives are enriched. Figure 21 shows the separation of the medians for the three docking models, -6.93 (actives) vs. -5.64 (inactives) for the PDE10A docking model, -7.66 (actives) vs. -6.01 (inactives) for the A_{2A}R docking model, and -7.60 (actives) vs. -5.66 (inactives) for the A₁R docking model. Statistical analysis was performed with R using a Mann-Whitney test[337] on the active and inactive docking score distributions of each target. The differences in medians were significant with p values less than 0.05 (Script [338]).

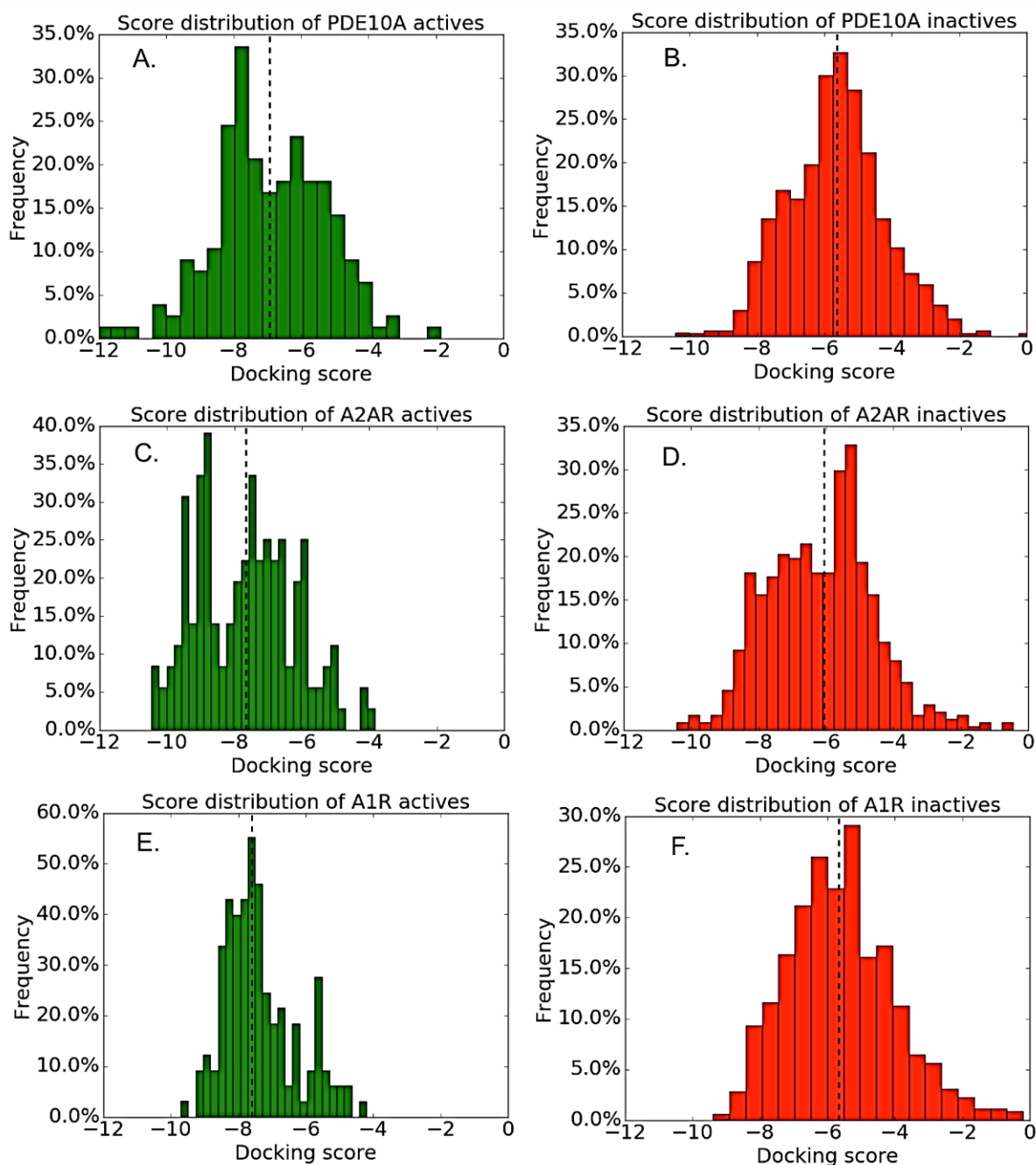


Figure 21. The separation in the medians (dashed lines) obtained for the three docking models **A.** -6.93 (ChEMBL actives) **B.** -5.64 (PubChem in-actives) for the PDE10A docking model, **C.** -7.66 (ChEMBL actives) **D.** -6.01 (PubChem inactives) for the A_{2A}R docking model, **E.** -7.60 (ChEMBL actives) **F.** -5.66 (PubChem inactives) for the A₁R docking model, indicating that actives are enriched in the three docking models

The F_1 score which is the harmonic mean of precision and recall, was computed (using a Python script[338]) for all the docking scores of the ChEMBL actives and PubChem inactives for each model. A search was performed for a docking score threshold that gave the highest F_1 score. Subsequently, substructure analysis was performed on compounds that were simultaneously predicted as A₁R/A_{2A}R-PDE10A multi-target ligands by target prediction, which also

displayed docking scores that are lower than or equal to those with the highest F_1 score for each of the three docking models (A_1R , $A_{2A}R$, and PDE10A, using script provided by Kalash *et al* [338]). Furthermore, the thresholds found are intended to serve as reference scores for any structure-based design problem at these target classes.

3.2.7 Docking

The RECAP compounds that were predicted as $A_1R/A_{2A}R$ -PDE10A multi-target ligands were docked against the $A_{2A}R$ protein crystal structure (PDB ID: 4EIY),[368] the A_1R homology model and the PDE10A protein crystal structure (PDB ID: 4DDL)[369] to use their predicted binding energies as a filter, and investigate the molecular interactions. The Glide docking parameters used here are given in Table 2. The parameters were deduced from docking experiments using known actives and inactives against each protein model.

Table 2. Glide docking parameters used for the A_1R , $A_{2A}R$, and PDE10A models were deduced from docking experiments using known actives and inactives against each protein model

Docking model	A_1R	$A_{2A}R$	PDE10A
Precision option	Standard precision (SP) ^a	Standard precision (SP) ^a	Extra precision (XP) ^a
Ligand sampling option	Flexible ^b	Flexible ^b	Flexible ^b

^a SP option allows better coverage of conformational space whereas XP option gives higher accuracy on docked poses ^b Flexible ligand sampling is a default choice, which generates conformations internally during the docking process

3.2.8 Substructural analysis

Subsequently, substructure analysis was performed using DataWarrior 4.2.2, on the proposed $A_1R/A_{2A}R$ -PDE10A multi-target ligands predicted by both ligand-based and structure-based techniques (considering docking scores less than or equal to the threshold of the best F measure for each docking model). The chemical series found were [1,2,4]triazolo[1,5-c]quinazolines (50.4%), imidazo[1,5-a]quinoxalines (14.4%), 6,7-alkoxyisoquinolines (10.6%), and 2-aminopyridine-3-carbonitriles (9.2%), which are depicted in Figure 22, in addition to various compounds consisting of the common and frequent heterocycles identified originally in the substructural analysis of the extracted ChEMBL compounds.

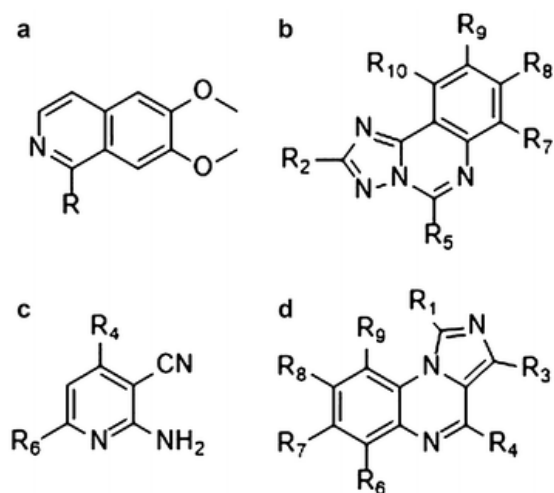


Figure 22. 2,563 compounds of the focused RECAP library were predicted as A₁R/A_{2A}R-PDE10A multi-target ligands, and docked against the A_{2A}R protein crystal structure (PDB ID: 4EIY), A₁R homology model, and the PDE10A protein crystal structure (PDB ID: 4DDL). The RECAP series which showed an agreement between the ligand-based and structure-based predictions were mainly **a.** 6,7-alkoxyisoquinolines **b.** [1,2,4]triazolo[1,5-c]quinazolines **c.** 2-aminopyridine-3-carbonitriles **d.** imidazo[1,5-a]quinoxalines

3.2.9 Experimental validation of 2-aminopyridine-3-carbonitriles as multi-target ligands at A₁R, A_{2A}R, and PDE10A

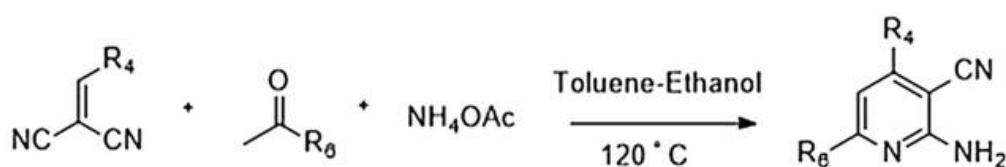
The experimental validation of 2-aminopyridine-3-carbonitriles as multi-target ligands at A₁R, A_{2A}R, and PDE10A, included the synthesis, characterization and pharmacological evaluation of the compounds **11–35**. Cristina Val (CV) and Jhonny Azuaje (JA) at the Center for Research in Biological Chemistry and Molecular Materials (CIQUS) in the University of Santiago de Compostela performed the experiments.

3.2.9.1 Synthesis of novel 4,6-substituted 2-amino-pyridin-3-carbonitriles

Due to both synthetic accessibility of the reaction and yield, CV optimized a one-pot synthetic scheme for the purpose of synthesizing 2-aminopyridine-3-carbonitriles. For the other series, the synthetic routes were multi-step reactions, which due to synthetic complexity were not considered for synthesis.

The synthetic routes reported in the literature for the formation of derivatives of 6,7-alkoxyisoquinolines as selective PDE10A inhibitors involved multi-step reactions ranging from 3 to 13 steps.[375, 376] Whereas, the procedures for the synthesis of the imidazo[1,5-a]quinoxalines, known PDE10A inhibitors, consisted of 3 to 7 step reactions.[377–379] The [1,2,4]triazolo[1,5-c]quinazolines have been reported as potent and selective A_{2A}R antagonists and PDE10A inhibitors, and their synthesis involved 4 to 7 step reactions.[380–382]

Hence, given the fact that the 2-aminopyridine-3-carbonitriles were the only RECAP series that could be synthesized *via* a one-pot synthetic scheme,[383–385] they were selected for synthesis and subsequent validation as multi-target ligands by CV and JA. In particular, the compounds selected for synthesis did not exhibit any potential PAINs liability upon screening with the FAFDrug3 ADME-Tox Filtering Tool.[313] The one-pot synthetic route is shown in Scheme 1.



Scheme 1. The one-pot synthetic route followed for the synthesis of novel 4,6-substituted 2-amino-pyridin-3-carbonitriles

Chemistry. Unless otherwise indicated, all starting materials, reagents and solvents were purchased and used without further purification. After extraction from aqueous phases, the organic solvents were dried over anhydrous sodium sulfate. The reactions were monitored by thin-layer chromatography (TLC) on 2.5 mm Merck silica gel GF 254 strips, and each of the purified compounds showed a single spot; unless stated otherwise, UV light and/or iodine vapor were used to detect compounds. The synthesis of the target compounds was performed in coated Kimble vials on a PLS (6×4) Organic Synthesizer with orbital stirring. Filtration and washing protocols for supported reagents were performed in a 12-channel vacuum manifold. The purity and identity of all tested compounds were established by a combination of HPLC, elemental analysis, mass spectrometry and NMR spectroscopy as described below. Purification of isolated products was carried out by column chromatography (Kieselgel 0.040–0.063 mm, E. Merck) or medium pressure liquid chromatography (MPLC) on a CombiFlash Companion (Teledyne ISCO) with RediSep pre-packed normal-phase silica gel (35–60 μm) columns followed by recrystallization. Melting points were determined on a Gallenkamp melting point apparatus and are uncorrected. The NMR spectra were recorded on Bruker AM300 and XM500 spectrometers. Chemical shifts are given as δ values against tetramethylsilane as internal standard and J values are given in Hz. Mass spectra were obtained on a Varian MAT-711 instrument. Analytical HPLC was performed on an Agilent 1100 system using an Agilent Zorbax SB-Phenyl, 2.1 mm × 150 mm, 5 μm column with gradient elution using the mobile phases (A) H₂O containing 0.1% CF₃COOH and (B) MeCN and a flow rate of 1 mL/min. The purity of all tested compounds was determined to be greater than or equal to 95% (work done by CV and JA).

The synthesis of the 4,6-substituted 2-amino-pyridin-3-carbonitriles **11-35** was done *via* the one-pot synthetic route shown in Scheme 1. Varying both substituents on the ylidene malononitrile and the ketone reagents resulted in a variation of the substituents on positions 4 and 6 of the pyridine ring.

Synthetic procedure

Substituted ylidene malononitrile (1.0mmol), ketone (1.0mmol) and ammonium acetate (5.0mmol) in a 1:1 toluene/EtOH mixture (7mL) were stirred in a coated Kimble vial at 120°C for 12-24 h. After reaction completion (TLC control), distilled water was added and the mixture was extracted with ethyl acetate (3 x 10 mL). The organic phase was dried (Na₂SO₄) and evaporated under reduced pressure to afford an oily residue that was purified by column chromatography using *n*-hexane – ethyl acetate in 2:1 mixture (work done by CV and JA).

2-amino-6-(4-fluorophenyl)-4-phenylpyridine-3-carbonitrile (11)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.246 g, 85% yield (97% purity by HPLC). MP 226–228 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 8.08–7.95 (m, 2H), 7.69–7.58 (m, 2H), 7.60–7.47 (m, 3H), 7.23–7.09 (m, 3H), 5.34 (s, 2H). MS (EI) *m/z* (%): 289.07 (M⁺, 100), 262.07 (7). Analysis calculated for C₁₈H₁₂FN₃: C, 74.73; H, 4.18; F, 6.57; N 14.52. Found: C, 74.70; H, 4.19; F, 6.55; N, 14.54.

2-amino-6-(4-hydroxyphenyl)-4-phenylpyridine-3-carbonitrile (12)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.227 g, 79% yield (96% purity by HPLC). MP 241–243 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 9.92 (s, 1H), 7.99 (d, *J* = 8.6 Hz, 2H), 7.78–7.59 (m, 2H), 7.58–7.47 (m, 3H), 7.15 (s, 1H), 6.88 (s, 2H), 6.83 (d, *J* = 8.7 Hz, 2H). MS (EI) *m/z* (%): 287.04 (M⁺, 100), 259.89 (10). Analysis calculated for C₁₈H₁₃N₃O: C, 75.25; H, 4.56; N, 14.63; O, 5.57. Found: C, 75.27; H, 4.54; N, 14.62; O, 5.59.

2-amino-4-phenyl-6-(1,3-thiazol-2-yl)pyridine-3-carbonitrile (13)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.172 g, 62% yield (95% purity by HPLC). MP 154–156 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 7.95 (d, *J* = 3.0 Hz, 1H), 7.72 (s, 1H), 7.66–7.65 (m, 2H), 7.52–7.50 (m, 4H), 5.30 (s, 2H). MS (EI) *m/z* (%): 278.03 (M⁺, 100), 276.97 (45). Analysis calculated for C₁₅H₁₀N₄S: C, 64.73; H, 3.62; N, 20.13; S, 11.52. Found: C, 64.85; H, 3.48; N, 20.25; S, 11.42.

2-amino-6-(1-methyl-1H-pyrrol-2-yl)-4-phenylpyridine-3-carbonitrile (14)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.189 g, 69% yield (98% purity by HPLC). MP 152–153 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 7.67–7.54 (m, 2H), 7.56–7.42 (m, 3H), 7.30 (s, 1H), 6.91 (s, 1H), 6.66–6.59 (m, 2H), 5.23 (s, 2H), 3.70 (s, 3H). MS (EI) *m/z* (%): 274.14 (M⁺, 100). Analysis calculated for C₁₇H₁₄N₄: C, 74.43; H, 5.14; N, 20.42. Found: C, 74.57; H, 5.12; N, 20.30.

2-amino-4-(2-methoxyphenyl)-6-phenylpyridine-3-carbonitrile (15)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.238 g, 79% yield (97% purity by HPLC). MP 199–200 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 8.03–7.93 (m, 2H), 7.52–7.41 (m, 4H), 7.31 (dd, J₁ = 7.5 Hz, J₂ = 1.8 Hz, 1H), 7.17 (s, 1H), 7.11–7.02 (m, 2H), 5.27 (s, 2H), 3.88 (s, 3H). MS (EI) *m/z* (%): 301.16 (M⁺, 100), 270.12 (7), 120.10 (16.3). Analysis calculated for C₁₉H₁₅N₃O: C, 75.73; H, 5.02; N, 13.94; O, 5.31. Found: C, 75.76; H, 5.04; N, 13.92; O, 5.33.

2-amino-4-(2,4-dimethoxyphenyl)-6-phenylpyridine-3-carbonitrile (16)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.238 g, 72% yield (99% purity by HPLC). MP 155–157 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 8.02–7.90 (m, 2H), 7.52–7.38 (m, 3H), 7.32–7.22 (m, 1H), 7.16 (s, 1H), 6.69–6.55 (m, 2H), 5.25 (s, 2H), 3.88 (s, 3H), 3.86 (s, 3H). MS (EI) *m/z* (%): 331.14 (M⁺, 100), 165.51 (9), 120.16 (11.3). Analysis calculated for C₂₀H₁₇N₃O₂: C, 72.49; H, 5.17; N, 12.68; O, 9.66. Found: C, 72.50; H, 5.19; N, 12.71; O, 9.70.

2-amino-4-(2H-1,3-benzodioxol-5-yl)-6-phenylpyridine-3-carbonitrile (17)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.236 g, 75% yield (96% purity by HPLC). MP 220–221 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 8.12–7.86 (m, 2H), 7.56–7.38 (m, 3H), 7.20–7.08 (m, 3H), 6.95 (d, J = 8.0 Hz, 1H), 6.06 (s, 2H), 5.33 (s, 2H). MS (EI) *m/z* (%): 315.11 (M⁺, 100), 157.52 (5). Analysis calculated for C₁₉H₁₃N₃O₂: C, 72.37; H, 4.16; N, 13.33; O, 10.15. Found: C, 72.45; H, 4.06; N, 13.49; O, 10.00.

2-amino-4-cyclohexyl-6-phenylpyridine-3-carbonitrile (18)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.216 g, 78% yield (98% purity by HPLC). MP 125–126 °C. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.95–7.92 (m, 1H), 7.53–7.43 (m, 3H), 7.05 (s, 1H), 6.73 (s, 1H), 5.22 (s, 2H), 2.90–2.85 (m, 2H), 1.90–1.78 (m, 4H), 1.52–1.39 (m, 4H), 1.33–1.25 (m, 1H). MS (EI) *m/z* (%): 277.25 (M⁺, 74), 246.15 (56), 222.15 (100). Analysis calculated for C₁₈H₁₉N₃: C, 77.95; H, 6.90; N, 15.15. Found: C, 78.03; H, 6.96; N, 15.01.

2-amino-4-cyclohexyl-6-(2-fluorophenyl)pyridine-3-carbonitrile (19)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.186 g, 63% yield (95% purity by HPLC). MP 126–127 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 7.89 (td, *J* = 7.8, 1.9 Hz, 1H), 7.47–7.31 (m, 1H), 7.25–7.03 (m, 3H), 5.18 (s, 2H), 2.98–2.67 (m, 1H), 1.99–1.73 (m, 5H), 1.53–1.16 (m, 5H). MS (EI) *m/z* (%): 295.15 (M⁺, 98.05), 263.05 (23.28), 251.00 (12), 240.00 (100). Analysis calculated for C₁₈H₁₈FN₃: C, 73.20; H, 6.14; F, 6.43; N, 14.23. Found: C, 73.22; H, 6.17; F, 6.44; N, 14.25.

2-amino-4-cyclohexyl-6-(2-methylphenyl)pyridine-3-carbonitrile (20)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.236 g, 81% yield (97% purity by HPLC). MP 120–121 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 7.73–7.10 (m, 4H), 6.71 (s, 1H), 5.20 (s, 2H), 2.95–2.77 (m, 1H), 2.35 (s, 3H), 2.01–1.69 (m, 5H), 1.56–1.34 (m, 4H), 1.34–1.18 (m, 1H). MS (EI) *m/z* (%): 291.14 (M⁺, 100), 236.12 (48), 208.10 (91.7). Analysis calculated for C₁₉H₂₁N₃: C, 78.32; H, 7.26; N, 14.42. Found: C, 78.48; H, 7.18; N, 14.34.

2-amino-4-cyclohexyl-6-(thiophen-2-yl)pyridine-3-carbonitrile (21)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.167 g, 59% yield (98% purity by HPLC). MP 160–162 °C. ¹H NMR (300 MHz, CDCl₃), δ(ppm) 7.63–7.62(m, 1H), 7.44 (d, *J* = 4.5 Hz, 1H), 7.12–7.09 (m, 1H), 6.96 (s, 1H), 5.14 (s, 2H), 2.82–2.79 (m, 1H), 1.90–1.78 (m, 5H), 1.55–1.43 (m, 4H), 1.30–1.19 (m, 1H). MS (EI) *m/z* (%): 283.04 (M⁺, 100), 251.99 (19), 228.02 (92). Analysis calculated for C₁₆H₁₇N₃S: C, 67.81; H, 6.05; N, 14.83; S, 11.31. Found: C, 67.89; H, 6.13; N, 14.77; S, 11.21.

2-amino-4-cyclohexyl-6-(thiophen-3-yl)pyridine-3-carbonitrile (22)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.147 g, 52% yield (96% purity by HPLC). MP 145–146 °C. ¹H NMR (300 MHz, CDCl₃), δ(ppm) 7.94 (dd, *J* = 3.0, 1.3 Hz, 1H), 7.59 (dd, *J* = 5.1, 1.3 Hz, 1H), 7.38 (dd, *J* = 5.1, 3.0 Hz, 1H), 6.93 (s, 1H), 5.14 (s, 2H), 2.95–2.73 (m, 1H), 2.06–1.73 (m, 5H), 1.56–1.37 (m, 4H), 1.38–1.19 (m, 1H). MS (EI) *m/z* (%): 283.07 (M⁺, 100), 228.04 (93), 214.96 (52). Analysis calculated for C₁₆H₁₇N₃S: C, 67.81; H, 6.05; N, 14.83; S, 11.31. Found: C, 67.91; H, 6.09; N, 14.67; S, 11.33.

2-amino-4-cyclohexyl-6-(furan-2-yl)pyridine-3-carbonitrile (23)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.174 g, 65% yield (98% purity by HPLC). MP 177–178 °C. ¹H NMR (300 MHz, CDCl₃), δ(ppm) 7.55 (dd, *J* = 1.7, 0.8 Hz, 1H), 7.06 (dd, *J* = 3.4, 0.8 Hz, 1H), 7.03 (s, 1H), 6.54 (dd, *J* = 3.5, 1.8 Hz, 1H), 5.15 (s, 2H), 3.01–2.68 (m, 1H), 2.04–1.74 (m, 5H), 1.55–1.39

(m, 4H), 1.34–1.20 (m, 1H). MS (EI) m/z (%): 267.11 (M^+ , 100), 212.02 (69). Analysis calculated for $C_{16}H_{17}N_3O$: C, 71.89; H, 6.41; N, 15.72; O, 5.98. Found: C, 71.91; H, 6.43; N, 15.71.

2-amino-6-(2-fluorophenyl)-4-(4-methoxyphenyl)pyridine-3-carbonitrile (24)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.188 g, 59% yield (97% purity by HPLC). MP 180–181 °C. 1H NMR (300 MHz, $CDCl_3$), δ (ppm) 7.96 (td, $J_1 = 7.8$, $J_2 = 1.9$ Hz, 1H), 7.65–7.58 (m, 2H), 7.47–7.37 (m, 1H), 7.31–7.23 (m, 2H), 7.23–7.09 (m, 1H), 7.09–6.98 (m, 2H), 5.32 (s, 2H), 3.88 (s, 3H). MS (EI) m/z (%): 319.12 (M^+ , 100), 304.18 (12), 249.13 (16). Analysis calculated for $C_{19}H_{14}FN_3O$: C, 71.46; H, 4.42; F, 5.95; N, 13.16; O, 5.01. Found: C, 71.48; H, 4.44; F, 5.97; O, 5.05.

2-amino-4-(4-methoxyphenyl)-6-(2-methylphenyl)pyridine-3-carbonitrile (25)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.205 g, 65% yield (95% purity by HPLC). MP 151–152 °C. 1H NMR (300 MHz, $CDCl_3$), δ (ppm) 7.61 (d, $J = 8.3$ Hz, 2H), 7.40 (d, $J = 7.3$ Hz, 1H), 7.37–7.27 (m, 3H), 7.03 (d, $J = 8.2$ Hz, 2H), 6.86 (s, 1H), 5.32 (s, 2H), 3.87 (s, 3H), 2.42 (s, 3H). MS (EI) m/z (%): 314.10 (M^+ , 100), 271.06 (7), 208.11 (52). Analysis calculated for $C_{20}H_{17}N_3O$: C, 76.17; H, 5.43; N, 13.32; O, 5.07. Found: C, 76.31; H, 5.33; N, 13.52; O, 4.84.

2-amino-6-(furan-2-yl)-4-(4-methoxyphenyl)pyridine-3-carbonitrile (26)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.198 g, 68% yield (99% purity by HPLC). MP 205–207 °C. 1H NMR (300 MHz, $CDCl_3$) δ (ppm): 7.65–7.54 (m, 3H), 7.16 (s, 1H), 7.11 (d, $J = 3.5$ Hz, 1H), 7.03 (d, $J = 8.8$ Hz, 2H), 6.62–6.51 (m, 1H), 5.30 (s, 2H), 3.88 (s, 3H). MS (EI) m/z (%): 291.12 (M^+ , 100), 145.63 (5). Analysis calculated for $C_{17}H_{13}N_3O_2$: C, 70.09; H, 4.50; N, 14.42; O, 10.98. Found: C, 70.21; H, 4.38; N, 14.68; O, 10.73.

2-amino-6-(4-hydroxyphenyl)-4-(4-methoxyphenyl)pyridine-3-carbonitrile (27)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.222 g, 70% yield (99% purity by HPLC). MP 248–250 °C. 1H NMR (300 MHz, $CDCl_3$), δ (ppm) 9.89 (s, 1H), 7.98 (d, $J = 8.7$ Hz, 2H), 7.61 (d, $J = 8.7$ Hz, 2H), 7.11–7.06 (m, 3H), 6.84–6.81 (m, 4H), 3.82 (s, 3H). MS (EI) m/z (%): 317.17 (M^+ , 100), 302.04 (6), 158.50 (14). Analysis calculated for $C_{19}H_{15}N_3O_2$: C, 71.91; H, 4.76; N, 13.24; O, 10.08. Found: C, 71.94; H, 4.79; N, 13.25; O, 10.11.

2-amino-4,6-bis(2-fluorophenyl)pyridine-3-carbonitrile (28)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.219 g, 73% yield (98% purity by HPLC). MP 180–181 °C. 1H NMR (300 MHz,

CDCl₃), δ (ppm) 8.05–7.90 (m, 1H), 7.56–7.41 (m, 2H), 7.33–7.06 (m, 6H), 5.34 (s, 2H). MS (EI) m/z (%): 307.06 (M^+ , 100), 279.99 (8). Analysis calculated for C₁₈H₁₁F₂N₃: C, 70.35; H, 3.61; F, 12.36; N, 13.67. Found: C, 70.37; H, 3.63; F, 12.33; N, 13.66.

2-amino-6-(2-fluorophenyl)-4-(2-methoxyphenyl)pyridine-3-carbonitrile (29)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.245 g, 78% yield (97% purity by HPLC). MP 187–188 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm) 7.97 (td, J = 7.8, 1.9 Hz, 1H), 7.52–7.35 (m, 2H), 7.31 (td, J = 7.2, 1.5 Hz, 1H), 7.26–7.19 (m, 2H), 7.17–6.95 (m, 3H), 5.27 (s, 2H), 3.88 (s, 3H). MS (EI) m/z (%): 319.12 (M^+ , 100), 290.14 (7), 138.01 (14). Analysis calculated for C₁₉H₁₄N₃FO: C, 71.46; H, 4.42; F, 5.95; N, 13.16; O, 5.01. Found: C, 71.44; H, 4.43; F, 5.92; O, 5.04.

2-amino-4-(2-methoxyphenyl)-6-(2-methylphenyl)pyridine-3-carbonitrile (30)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.186 g, 64% yield (98% purity by HPLC). MP 181–183 °C. ¹H NMR (300 MHz, CDCl₃) δ (ppm): 7.47–7.40 (m, 2H), 7.32–7.28 (m, 4H), 7.09–7.02 (m, 2H), 6.86 (s, 1H), 5.29 (s, 2H), 3.88 (s, 3H), 2.43 (s, 3H). MS (EI) m/z (%): 315.13 (M^+ , 100), 298.16 (12), 284.09 (18), 208.10 (81.6). Analysis calculated for C₂₀H₁₇N₃O: C, 76.17; H, 5.43; N, 13.32; O, 5.07. Found: C, 76.19; H, 5.41; N, 13.36; O, 5.03.

2-amino-6-(furan-2-yl)-4-(2-methoxyphenyl)pyridine-3-carbonitrile (31)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.244 g, 77% yield (96% purity by HPLC). MP 187–188 °C. ¹H NMR (300 MHz, CDCl₃), δ (ppm): 7.55 (s, 1H), 7.44 (t, J = 8.1 Hz, 1H), 7.30 (dd, J = 7.4, 1.7 Hz, 1H), 7.15–6.98 (m, 4H), 6.54 (dd, J = 3.3, 1.7 Hz, 1H), 5.24 (s, 2H), 3.87 (s, 3H). MS (EI) m/z (%): 291.10 (M^+ , 100), 262.14 (10). Analysis calculated for C₁₇H₁₃N₃O₂: C, 70.09; H, 4.50; N, 14.42; O, 10.98. Found: C, 70.11; H, 4.51; N, 14.41; O, 11.01.

2-amino-6-(4-hydroxyphenyl)-4-(2-methoxyphenyl)pyridine-3-carbonitrile (32)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.193 g, 60% yield (96% purity by HPLC). MP 210–212 °C. ¹H NMR (300 MHz, DMSO-*d*₆), δ (ppm): 9.91 (s, 1H), 7.93 (d, J = 9.0 Hz, 2H), 7.45 (t, J = 7.8 Hz, 1H), 7.29 (dd, J = 7.4, 1.7 Hz, 1H), 7.16 (d, J = 8.3 Hz, 1H), 7.07 (d, J = 7.5 Hz, 1H), 7.03 (s, 1H), 6.82 (d, J = 8.9 Hz, 2H), 6.77 (s, 2H), 3.77 (s, 3H). MS (EI) m/z (%): 317.13 (M^+ , 100), 300.09 (8), 286.11 (6). Analysis calculated for C₁₉H₁₅N₃O₂: C, 71.91; H, 4.76; Cl, 13.24; O, 10.08. Found: C, 71.92; H, 4.74; Cl, 13.27; O, 10.05.

2-amino-4-(2-chlorophenyl)-6-(4-hydroxyphenyl)pyridine-3-carbonitrile (33)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.179 g, 59% yield (98% purity by HPLC). MP 215–217 °C. ¹H NMR (300 MHz, DMSO-*d*₆), δ (ppm): 9.90 (s, 1H), 8.16–7.22 (m, 2H), 7.69–7.30 (m, 4H), 7.16–6.50 (m, 5H). MS (EI) *m/z* (%): 320.99 (M⁺, 100), 286.04 (5). Analysis calculated for C₁₈H₁₂ClN₃O: C, 67.19; H, 3.76; Cl, 11.02; N, 13.06; O, 4.97. Found: C, 67.37; H, 3.94; Cl, 11.18; N, 12.88; O, 4.63.

2-amino-4,6-bis(4-hydroxyphenyl)pyridine-3-carbo-nitrile (34)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.151 g, 53% yield (97% purity by HPLC). MP 299–300 °C. ¹H NMR (300 MHz, DMSO-*d*₆), δ (ppm) 9.92 (s, 2H), 8.19–7.79 (m, 2H), 7.68–7.37 (m, 2H), 7.42–6.99 (m, 1H), 7.01–6.62 (m, 6H). MS (EI) *m/z* (%): 303.06 (M⁺, 100), 184.01 (6). Analysis calculated for C₁₈H₁₃N₃O₂: C, 71.28; H, 4.32; N, 13.85; O, 10.55. Found: C, 71.40; H, 4.54; N, 13.75; O, 10.31.

2-amino-4-(furan-2-yl)-6-(thiophen-3-yl)pyridine-3-carbonitrile (35)

Purified by column chromatography (*n*-hexane-ethyl acetate 2:1) and then recrystallized from EtOH to give 0.123 g, 46% yield (95% purity by HPLC). MP 156–157 °C. ¹H NMR (300 MHz, CDCl₃) δ(ppm): 8.01 (dd, *J* = 3.0, 1.2 Hz, 1H), 7.66 (dd, *J* = 5.1, 1.2 Hz, 1H), 7.62 (dd, *J* = 1.8, 0.6 Hz, 1H), 7.48 (dd, *J* = 3.6, 0.6 Hz, 1H), 7.45 (s, 1H), 7.40 (dd, *J* = 5.1, 3.0 Hz, 1H), 7.40 (dd, *J* = 5.1, 3.0 Hz, 1H), 6.61 (dd, *J* = 3.6, 1.8 Hz, 1H), 5.26 (s, 2H). MS (EI) *m/z* (%): 267.06 (M⁺, 100), 237.98 (6), 210.99 (7). Analysis calculated for C₁₄H₉N₃OS: C, 62.91; H, 3.39; N, 15.72; O, 5.99; S, 11.99. Found: C, 63.11; H, 3.47; N, 15.58; O, 5.97; S, 11.87.

3.2.9.2 Pharmacological evaluation of novel 4,6-substituted 2-amino-pyridin-3-carbonitriles

Pharmacological evaluation was performed by CV. The assay used a radioligand binding competition assay, with A₁, A_{2A}, A_{2B}, and A₃ human receptors expressed in transfected CHO (A₁), HeLa (A_{2A} and A₃), and HEK-293 (A_{2B}) according to the procedure reported by Bosch *et al.*[385]

The activity measurements against the phosphodiesterases PDE7A, PDE7B, PDE9A and PDE10A were performed using AD293 cells that were transiently and separately transfected with human PDE7A, PDE7B, PDE9A, and PDE10A following the procedure described by Shipe *et al.*[386] The IC₅₀ values were obtained by fitting the data with non-linear regression

using Prism 2.1 software (GraphPad, San Diego, CA),[387] and the reported results are the mean of three experiments (n = 3) each performed in duplicate.

3.3 Results and Discussion

3.3.1 Design of synthetically feasible A₁R/A_{2A}R-PDE10A multi-target ligands

Human enzyme and receptor data were extracted from ChEMBL 20.[171] Substructure analysis of A₁R, A_{2A}R ligands and PDE10A inhibitors with K_i and IC₅₀ values less than or equal to 1 μ M revealed that the most frequently occurring common heterocycles among the actives against the three target classes were pyridine, pyrimidine, piperazine, and 1H-pyrazole (Figure 19). Subsequently, A₁R (2,104), A_{2A}R (2,489) and PDE10A inhibitors (679) containing those frequent heterocycles were subjected to RECAP analysis/synthesis in MOE (see Methods for details). As a result, 458,839 (potentially) synthetically accessible ligands were found *in silico*. This list of candidates was filtered to those retaining the common heterocycles (listed above), in order to create a focused chemical space characteristic of A₁R, A_{2A}R and PDE10A (with the simultaneous trade-off of reduced novelty), giving rise to 22,233 compounds.

3.3.2 Target prediction of the designed RECAP library

To assess the likelihood of active compounds against A₁R, A_{2A}R and PDE10A *via* a complementary approach, PIDGIN 1.0 (Prediction including Inactivity), a tool which uses ECFP 4 circular Morgan fingerprints trained on ChEMBL actives and PubChem inactives was used to perform *in silico* target prediction for the focused RECAP library (22,233 compounds).[133] Subsequent enrichment analysis of the predictions was done using an estimation score, the average ratio as developed by Liggi *et al*[366] and *via* a Chi-square test. For targets to be considered as enriched according to these methods, the estimation score and the Chi-square test p value should be less than or equal to 0.01 and 0.05, respectively. Hence, upon analyzing the enrichment parameters for the A₁R, A_{2A}R and PDE10A targets that were predicted for the focused RECAP library (Figure 20), the three targets were predicted with an estimation score equal to 0 (enriched) as well as average ratios less than 0.1 (enriched) with Chi-squared p values less than 0.005. The percentage of RECAP compounds of the focused library that were predicted as actives against the A₁R, A_{2A}R and PDE10A targets were 51.1%, 52.8%, and 24.5% respectively. These numbers are relatively high, which however is understandable given that the input to the RECAP analysis consisted of experimentally established known ligands of the above protein targets.

3.3.3 Docking of the compounds predicted as A₁R/A_{2A}R-PDE10A multi-target ligands

In the next step, docking and further substructure analysis were performed on compounds of the focused RECAP library, which were predicted as A₁R/A_{2A}R-PDE10A multi-target ligands from the ligand-based side in the previous step. 2,563 compounds were predicted as actives against the three desired targets and they were subsequently docked against a high resolution (1.8Å) A_{2A}R protein crystal structure (PDB ID: 4EIY)[368], its corresponding A₁R homology model (see Methods for details) and PDE10A (PDB ID: 4DDL).[369]

Compounds, which were carried forward to substructural analysis, were selected when their docking score gave a value less than a pre-determined cut-off value computed from the docking scores. This cut-off value was evaluated as the docking score with the best F measure statistic obtained by docking a set of known actives and inactives against the protein crystal structures and the homology model (see Methods for details).

As a result, a distribution of RECAP compounds that were favorable as multi-target ligands by target prediction and docking was obtained, where 62.47% of the RECAP compounds that were predicted as A₁R/A_{2A}R-PDE10A multi-target ligands and docked against PDE10A exhibited docking scores lower than -6.49 (the threshold of the best F measure discriminating between actives and inactives for known ligands). Out of the RECAP compounds, which displayed docking scores, lower than -6.49 against PDE10A, 48.89% and 35.23% displayed docking scores lower than -7.26 and -8.49 against A₁R and A_{2A}R (the thresholds of the best F measures).

3.3.4 Substructure analysis of the compounds predicted as A₁R/A_{2A}R-PDE10A multi-target ligands

Substructure analysis was performed on compounds having a favorable assessment by target prediction and docking (i.e. those compounds whose docking scores were below the threshold for all three targets). The analysis revealed frequently occurring series, which shared the same core structure, and are shown in Figure 22.

The chemical series were identified as [1,2,4]triazolo[1,5-c]quinazolines (50.4% of all positively predicted multi-target ligands by *in silico* target prediction as well as docking), imidazo[1,5-a]quinoxalines (14.4%), 6,7-alkoxyisoquinolines (10.6%), and 2-aminopyridine-3-carbonitriles (9.2%). These were in addition to various compounds containing the common and frequent heterocycles identified earlier (15.4%). Each series identified could be considered for synthesis, SAR studies and validation as A₁R/A_{2A}R-PDE10A multi-target ligands.

3.3.5 Synthesis of novel 2-aminopyridine-3-carbonitriles

Due to both ease of the reaction and anticipated yield, a one-pot synthetic scheme was selected for synthesizing one promising series, 2-aminopyridine-3-carbonitriles. The design resulted in 25 compounds for synthesis of which 21 were novel compounds and four (**11**, **12**, **15**, and **27**) have previously been reported in the literature.[388–391] Compounds **11–35** were synthesized according to the synthetic route illustrated in Scheme 1, and all products were obtained with good yields, ranging from 46% to 85% (see Methods for details).

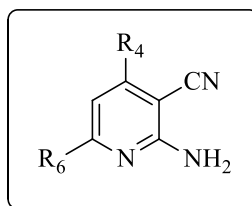
3.3.6 Pharmacological evaluation of novel 2-aminopyridine-3-carbonitriles

Bioactivity testing was performed using A₁ and A_{2A} human adenosine receptors expressed in transfected CHO (A₁) and HeLa (A_{2A}) cells, as well as AD293 cells that were transiently transfected with human PDE10A. Table 3 includes the list of synthesized 4,6-substituted 2-amino-pyridin-3-carbonitriles, along with their K_i values against A₁R, A_{2A}R, and IC₅₀ values against PDE10A. It can be seen that 15 compounds of the 25 synthesized 2-amino-pyridin-3-carbonitriles exhibited inhibitory activity against PDE10A below 10 μM. In addition, 13 compounds were adenosine receptor binders exhibiting selectivity towards A₁R and A_{2A}R, which is a new profile not seen in previous work reported by Mantri *et al.*, where 2-amino-pyridin-3-carbonitriles were promiscuous towards the four adenosine receptor subtypes.[390]

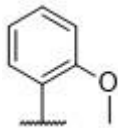
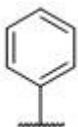
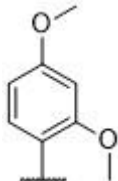
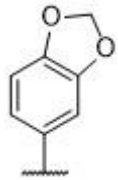

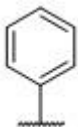
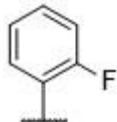
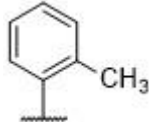
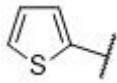
Given that the objective of this work is to find compounds displaying specific multi-target activity, compounds **18**, **26**, **31**, and **35** were identified as A₁R/A_{2A}R-PDE10A multi-target ligands, inhibiting PDE10A with IC₅₀ values of 2.4, 3.2, 10.0, and 5.1 μM respectively, and binding to A₁R with K_i values of 294 and 34 nM (compounds **18** and **26**, respectively), and to A_{2A}R with K_i values of 41, 95, and 55 nM (compounds **26**, **31**, and **35**, respectively). Notably, compound **26** exhibited the desired multi-target profile as a PDE10A inhibitor and a dual binder to A_{2A}R and A₁R.


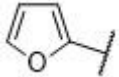
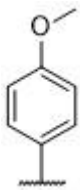
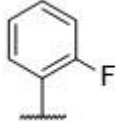
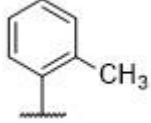
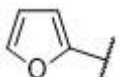

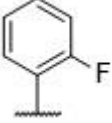
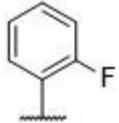
It was previously reported that substituted pyridines exhibited PDE inhibitory activity,[392, 393] and 2-amino-pyridin-3-carbonitriles are adenosine receptor ligands.[390] In this study suitable compounds matching *both* criteria as A₁R/A_{2A}R-PDE10A multi-target ligands, satisfying the original compound design objective have now been identified.

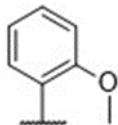
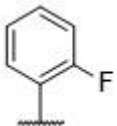
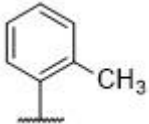
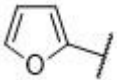
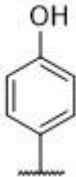
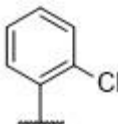
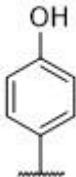
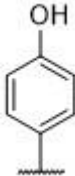
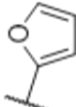

Table 3. Percent inhibition of the synthesized 4,6-substituted 2-amino-pyridin-3-carbonitriles at 10 μ M (PDE10A) or IC_{50} (μ M)^a and percentage displacement at 0.1 μ M (A_1R and $A_{2A}R$), or K_i ^b



Compound	R_4	R_6	% Inhibition at 10 μ M (PDE10A) or IC_{50} (μ M) % displacement at 0.1 μ M (A_1R and $A_{2A}R$) or K_i		
			A_1R	$A_{2A}R$	PDE10A
11			394 \pm 12 nM	32%	22%
12			142 \pm 7 nM	38%	52%
13			12%	8%	2.0 \pm 0.2 μ M
14			26%	32%	3.6 \pm 0.3 μ M

15			53%	543 ± 13 nM	28%
16			12%	1%	5.7 ± 0.3 μ M
17			25 ± 2 nM	5%	17%
18			294 ± 10 nM	50%	2.4 ± 0.2 μ M
19			84 ± 8 nM	34%	68%
20			17%	18%	3.7 ± 0.3 μ M
21			16%	11%	1.2 ± 0.1 μ M

22			44%	60%	$0.9 \pm 0.2 \mu\text{M}$
23			$70 \pm 3 \text{ nM}$	$49 \pm 4 \text{ nM}$	55%
24			$108 \pm 6 \text{ nM}$	30%	10%
25			6%	32%	$1.5 \pm 0.2 \mu\text{M}$
26			$34 \text{ nM} \pm 2 \text{ nM}$	$41 \pm 2 \text{ nM}$	$3.2 \pm 0.4 \mu\text{M}$
27			46%	29%	65%
28			$78 \pm 5 \text{ nM}$	$948 \pm 13 \text{ nM}$	38%

29			58%	338 ± 12 nM	73%
30			12%	50%	6.4 ± 0.4 μ M
31			38%	95 ± 4 nM	10.0 ± 0.6 μ M
32			8%	1%	5.6 ± 0.5 μ M
33			2%	10%	4.0 ± 0.3 μ M
34			19%	7%	3.1 ± 0.4 μ M
35			15%	55 ± 2 nM	5.1 ± 0.4 μ M

(a) IC₅₀ values of the 2-aminopyridines-3-carbonitriles were measured for the four phosphodiesterases PDE7A, PDE7B, PDE9A and PDE10A at 10 μ M concentration. For those compounds that showed percentage inhibition greater than 70% and selectivity against other

measured isoenzymes, IC_{50} were determined (b) Calculation of the K_i values at A_1R , $A_{2A}R$, $A_{2B}R$ and A_3R was approximated using the Cheng-Prusoff equation: $K_i = IC_{50}/[1 + (C/K_D)]$, where IC_{50} is the concentration of compound that displaces the binding of the radioligand by 50%, C is the concentration of radioligand, and K_D is the dissociation constant of each radioligand.

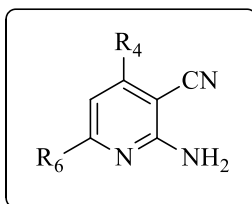
3.3.7 (SAR) Structure-activity relationship analysis

The purpose of the SAR analysis was to rationalize the variation in activity of the newly discovered $A_1R/A_{2A}R$ -PDE10A multi-target ligands against PDE10A, given that 2-amino-pyridin-3-carbonitriles are a novel class of PDE10A inhibitors. Compounds of this substructural class were also previously documented as adenosine receptor ligands.[390] Computational SAR studies focused on the PDE10A data, where the variation in potency was rationalized in relation to the physicochemical properties of the compounds (which were computed by FAFDrug3, Table 4).

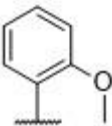
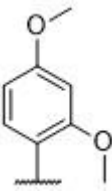

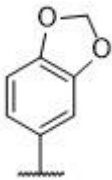


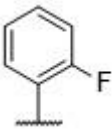
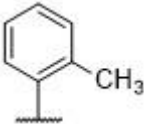
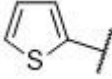
A trend observed repeatedly in several cases was that when logP decreased, associated with an increase in tPSA, then this led to an improvement in the activity against PDE10A. Initial analysis concentrated on compounds **11-14**, which have a phenyl substituent at position 4 of the pyridine ring. Compound **13** was the most potent PDE10A inhibitor with an IC_{50} of 2.0 μM , and a computed logP of 3.1 and tPSA of 103.9 \AA^2 . Similarly, for compounds **15-17** having a phenyl substituent at position 6 of the pyridine ring, compound **16** was the most potent against PDE10A with an IC_{50} of 5.7 μM and a computed logP of 4.0 and tPSA of 81.2 \AA^2 . For compounds **18-23**, which have a cyclohexyl ring at position 4 of the pyridine ring, compound **22** displayed the most potent PDE10A inhibitory activity with an IC_{50} of 0.9 μM and a computed logP of 4.7 and tPSA of 90.9 \AA^2 . For compounds **24-27**, with a p-methoxyphenyl substituent at position 4 of the pyridine ring, compound **26** with the smallest predicted lipophilicity of 3.1 and tPSA of 85.1 \AA^2 displayed a good PDE10A inhibitory activity with an IC_{50} value equal to 3.2 μM . A more potent compound was **25** with an IC_{50} value of 1.5 μM and a computed logP of 4.4 and tPSA of 71.9 \AA^2 . For compounds **29-32**, with an o-methoxyphenyl substituent at position 4 of the pyridine ring, compound **32** displayed PDE10A inhibitory activity with a higher potency (IC_{50} value of 5.6 μM), and a computed logP of 3.7 and tPSA of 92.2 \AA^2 . Finally, a similar general trend is observed for the compounds **33** and **34** with a 4-hydroxyphenyl substituent at position 6 of the pyridine ring, where compound **34** was a more potent PDE10A inhibitor with an IC_{50} of 3.1 μM and computed logP of 3.4 and tPSA of 103.2 \AA^2 . Hence, it could be deduced that in the majority of the series considered, where the


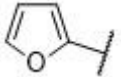
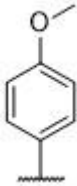
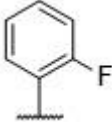
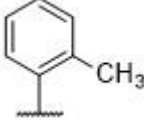
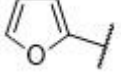

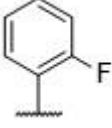
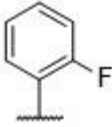
substituents on a single position is varied, a decrease in computed lipophilicity associated with an increase in polarity generally improved the activity of compounds against PDE10A. This general trend can be attributed to the hydrophilic nature of the pocket, which favors the interactions between the ligand and the PDE10A protein by compounds exhibiting these properties.

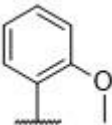
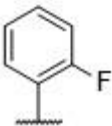
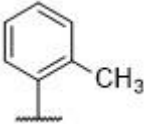
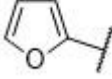

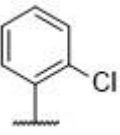
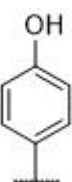

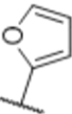

Table 4. Physicochemical properties (LogP and tPSA) computed for the synthesized 4,6-substituted 2-amino-pyridin-3-carbonitriles using FAFDrug3 ADME-Tox tool



Compound	R ₄	R ₆	logP	tPSA (Å ²)
11			4.2	62.7
12			3.7	82.9
13			3.1	103.8
14			2.8	67.6

15			4.0	71.9
16			4.0	81.2
17			3.9	81.2
18			5.0	62.7
19			5.1	62.7
20			5.3	62.7
21			4.7	90.9

22			4.7	90.9
23			4.1	75.8
24			4.1	71.9
25			4.4	71.9
26			3.1	85.1
27			3.6	92.2
28			4.3	62.7

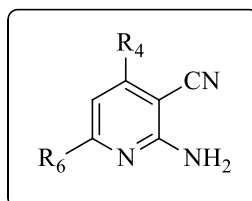
29			4.3	62.7
30			4.4	71.9
31			3.1	85.1
32			3.7	92.2
33			4.3	82.9
34			3.4	103.2
35			2.8	104.1

3.3.8 Compound selectivity assessment

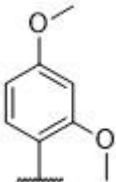
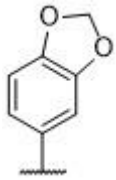

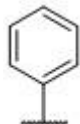
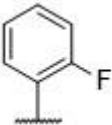
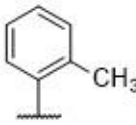
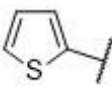

The selectivity of compounds **11-35** against the selected major off-targets A_{2B}R, A₃R, PDE7A, PDE7B, and PDE9A, was predicted using PIDGIN at a threshold for binding greater than or equal to 0.8, and subsequently tested experimentally. The IC₅₀ values were determined for compounds with % inhibition at phosphodiesterases greater than 70%. As shown in Table 5, the synthesized compounds are mostly inactive against those off-targets except for compounds **26**, **27**, **31**, and **33** that exhibited IC₅₀ values of 3.4, 3.5, 15.1 and 1.8 μ M against PDE7A, and compounds **33** and **35**, which exhibited IC₅₀ values of 7.3 and 4.7 μ M against PDE7B. Remarkably, compound **18** was found to exhibit selectivity over all tested off-targets using the above criterion, with the lowest selectivity measured for PDE7B (of 55% inhibition at 10 μ M ligand concentration). This can be compared to the IC₅₀ value of **18** at PDE10A, which is 2.4 μ M (indicating approximately 2-fold selectivity for **18**).

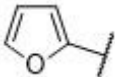
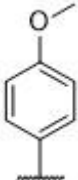
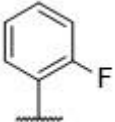
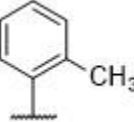
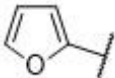
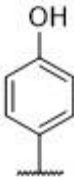
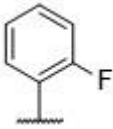
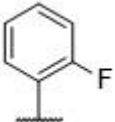
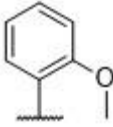
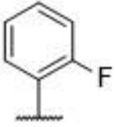
In general, the experimental results on off-targets for the synthesized 4,6-substituted 2-amino-pyridin-3-carbonitriles **11-35** agree with the predictions generated using PIDGIN utilized to bias the compound design towards more selective compounds such as **18** (Table 5). This compound would serve as a good starting point for analog modification to improve selectivity of the synthesized ligands towards PDE10A.

Table 5. Percentage inhibition of the synthesized 4,6-substituted 2-amino-pyridin-3-carbonitriles at 10 μ M (PDE7A, PDE7B, PDE9A), or IC₅₀ (μ M) and percentage displacement at 0.1 μ M (A_{2B}R and A₃R), or K_i



Compound	R ₄	R ₆	A _{2B} R	A ₃ R	PDE7A	PDE7B	PDE9A
11			2%	27%	9%	28%	1%
12			2%	2%	47%	39%	1%
13			3%	6%	24%	36%	5%
14			5%	13%	50%	48%	2%
15			5%	26%	24%	30%	3%

16			1%	6%	51%	27%	1%
17			1%	2%	22%	8%	19%
18			1%	13%	26%	55%	30%
19			2%	7%	19%	34%	1%
20			1%	28%	31%	35%	9%
21			1%	26%	33%	45%	6%
22			2%	14%	29%	60%	3%

23			3%	19%	44%	41%	4%
24			1%	27%	22%	20%	7%
25			4%	17%	57%	55%	16%
26			33%	21%	3.4 ± 0.4 μM	30%	2%
27			1%	8%	3.5 ± 0.4 μM	44%	7%
28			1%	1%	25%	36%	0%
29			2%	1%	49%	42%	19%

30			3%	6%	31%	23%	12%
31			2%	23%	15.1 ± 0.6 μM	30%	17%
32			1%	2%	63%	16%	9%
33			8%	1%	1.8 ± 0.3 μM	7.3 ± 0.3 μM	25%
34			1%	1%	47%	38%	7%
35			5%	2%	48%	4.7 ± 0.4 μM	14%

3.3.9 Analysis of the predicted binding modes of the synthesized 2-aminopyridine-3-carbonitriles

The synthesized 2-aminopyridine-3-carbonitriles were docked against A_{2A}R (PDB ID: 4EIY), the A₁R homology model, and PDE10A (PDB ID: 4DDL). Figure 23 shows the common predicted ligand-target interactions for representative multi-target ligands of A₁R-PDE10A, A₁R-A_{2A}R, and A_{2A}R-PDE10A, namely for compounds **18**, **28**, and **35**.

It can be seen that compounds **18** and **35**, with IC₅₀ values of 2.4 and 5.1 μ M respectively, share similarities in predicted binding modes, since their pyridine rings display π -stacking with Phe₆₈₆ and Phe₇₁₉ of PDE10A (Figure 23). These are the type of interactions predicted to be exhibited by the majority of the synthesized ligands from this work, as well as the only existing interactions between co-crystallized PDE10A inhibitors discovered by fragment screening (PDB IDs: 5C2E, 5C1W, 5C29, 5C2A having ligands with K_i values of 2, 8,700, 880, and 4.8 nM, respectively).[386] It is noted that the ligand bound to 5C2A exhibits a considerable selectivity towards PDE10A over all the other PDEs (in the range of 100-1000 fold and greater over the majority of PDEs, with the least selectivity observed being in the range of 25-100 fold). This ligand exhibits only π -stacking interactions with Phe₆₈₆ and Phe₇₁₉, similar to the mode of interactions of compound **18** with PDE10A, which is relatively selective over all tested PDEs, with the lowest selectivity being measured for PDE7B (of 55% inhibition at 10 μ M ligand concentration) and compound **35**, which is selective against all tested PDEs except PDE7B (Tables 3 and 5). Additional interactions were seen in analogs discovered by fragment screening, namely hydrogen bonding with Gln₇₁₆ and Tyr₆₈₃ in the PDE10A selectivity pocket (PDB IDs: 5C28 and 5C2H with K_i values of 2200 and 0.0082 nM respectively).[386] The ligand bound to 5C2H exhibits π -stacking with Phe₆₈₆ and Phe₇₁₉ and hydrogen bonding with Tyr₆₈₃ in the PDE10A selectivity pocket. The 5C2H ligand showed a very high selectivity towards PDE10A, greater than 5000-fold, which emphasizes the consideration of compound **18** for analog modification to target the selectivity pocket in order to improve the selectivity towards PDE10A. In addition, hydrogen bonding with Tyr₆₈₃ in the PDE10A selectivity pocket is also seen in many other highly selective PDE10A inhibitors reported in the literature[394] (PDB IDs: 5DH5,[395] 5B4L,[396] with K_i = 0.23 nM, and IC₅₀ = 0.76 nM respectively). This further highlights the importance of analog modification to target the PDE10A selectivity pocket.

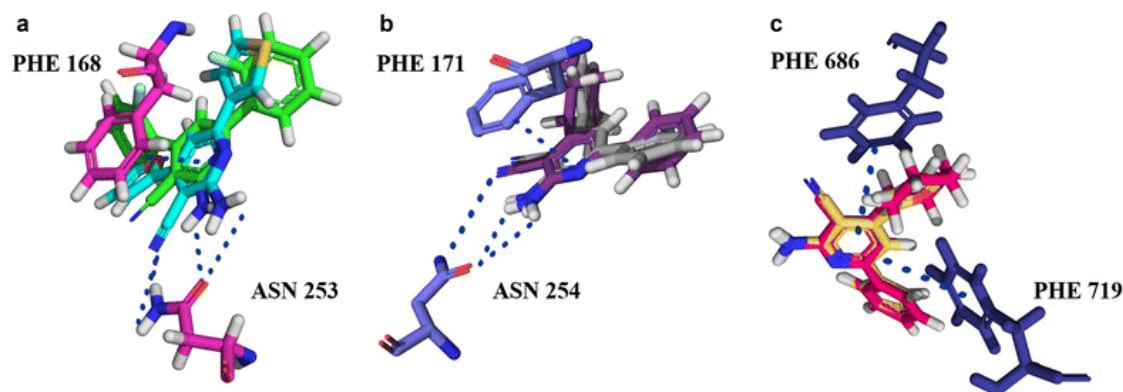


Figure 23. Docking studies predicted molecular interactions characteristic of the 4,6-substituted 2-amino-pyridin-3-carbonitriles with the A_{2A}R protein crystal structure (PDB ID: 4EIY), A₁R homology model, and PDE10A protein crystal structure (PDB ID: 4DDL), which are displayed for representative multi-target ligands with the following combinations: compound **18** (A₁R-PDE10A), **28** (A₁R-A_{2A}R), and **35** (A_{2A}R-PDE10A) **a.** interactions with A_{2A}R-the overlaid compounds **28** and **35** exhibit H-bonds *via* amino and carbonitrile groups with Asn₂₅₃, and the pyridine rings are π -stacked with Phe₁₆₈ **b.** interactions with A₁R-the overlaid compounds **18** and **28** exhibit H-bonds *via* amino and carbonitrile groups with Asn₂₅₄, and the pyridine rings are π -stacked with Phe₁₇₁ **c.** interactions with PDE10A-the overlaid compounds **18** and **35** have the pyridine rings π -stacked with Phe₆₈₆ and Phe₇₁₉. The molecular interactions predicted for the active molecules are consistent with observed interactions between co-crystallized ligands and their corresponding protein crystal structures (PDB IDs: 4EIY and 4DDL) and the interactions with the A₁R homology model reported in the literature.

Moreover, it is noted that compounds **26** and **31** with IC₅₀ values of 3.2 and 10.0 μ M respectively (which are selective against all tested PDEs except PDE7A, Tables 3 and 5) were predicted to exhibit an additional type of interaction, H-bonding with Gln₇₁₆ *via* their overlaid furan rings at position 6 of the pyridine ring (Figure 24). In fact H-bonding with Gln₇₁₆ was the only interaction, besides π -stacking with Phe₆₈₆ and Phe₇₁₉, which has been observed in many of the highly selective PDE10A ligands reported in the literature (PDB IDs: 4DDL,[369] 3SN7, 3SNL, and 3SNI,[379] 5DH4 and 5DH5[395] with IC₅₀ values of 4.9, 0.7, 0.7, 11 nM and K_i = 0.23 nM respectively). As for other type of interactions generally exhibited by known PDE10A inhibitors such as hydrogen bonding with Gln₇₂₆ and π -stacking with Phe₇₂₉ (PDB ID: 5EDE), [397]none were predicted for any of the compounds presented in this work.

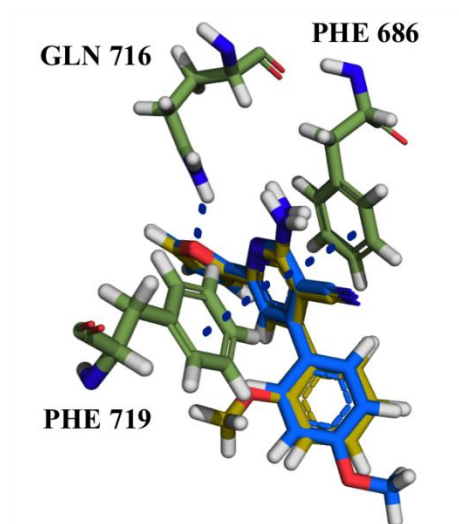


Figure 24. Docking studies predicted molecular interactions for the overlaid compounds **26** and **31** with PDE10A: π -stacking of the pyridine rings with Phe₆₈₆ and Phe₇₁₉, and H-bonding with Gln₇₁₆ *via* their overlaid furan rings at position 6 of the pyridine ring

Common predicted binding modes could also be observed for the synthesized compounds against the adenosine receptors A_{2A}R and A₁R. Figure 23 displays the interactions of two representative compounds **28** and **35**, which exhibit K_i values of 948 and 55 nM respectively. These are H-bonding of their pyridine rings with Asn₂₅₃ and π -stacking of their amino and carbonitrile groups with Phe₁₆₈ of A_{2A}R. As for A₁R, the overlaid compounds **18** and **28**, with K_i values of 294 and 78 nM respectively, H-bond *via* their amino and carbonitrile groups with Asn₂₅₄, and their pyridine rings are π -stacked against Phe₁₇₁. It can be observed that the ligand/protein interactions predicted for the active compounds against the A_{2A}R are also those seen in the co-crystallized ligand/protein crystal structures (PDB IDs: 4EIY,[368] 3EML,[398] 5IU4,[399] with a K_i value of 0.8 nM for ZM 241385, which is the ligand common to the three PDB IDs). The interactions with the A₁R are similar to those reported with the A₁R homology model (for ligands with IC₅₀ values of 2.9 and 6.2 nM (from the literature) and predicted to interact with the A₁R homology model).[400, 401]

Generally, the compounds exhibited good selectivity towards A₁R and A_{2A}R (Tables 3 and 5) with a nanomolar range of binding affinities. As for the selectivity towards PDE10A, this could be improved by analog modification of compound **18** to favor the hydrogen bonding interaction with Tyr₆₈₃ in the PDE10A selectivity pocket. In addition, the potency of compounds against PDE10A could also be optimized to achieve therapeutically relevant efficacy.

3.3.10 Computational assessment of CNS permeability

Compounds **18** and **26** exhibited the desired multi-target profile by inhibiting PDE10A and binding to A_{2A}R and/or A₁R. The physicochemical properties of these compounds were calculated by FAFDrug3,[313] and both compounds passed Lipinski rule of 5 and the CNS filter, which takes into consideration the assessment of their ability to pass the blood brain barrier (Figure 25).[304] Hence, while further experimental work would be needed to establish the validity of those predictions, compounds **18** and **26** may serve as good starting points for further functional efficacy assessment and selectivity optimization towards PDE10A, A_{2A}R and/or A₁R for subsequent consideration of development for the treatment of neurodegenerative diseases.

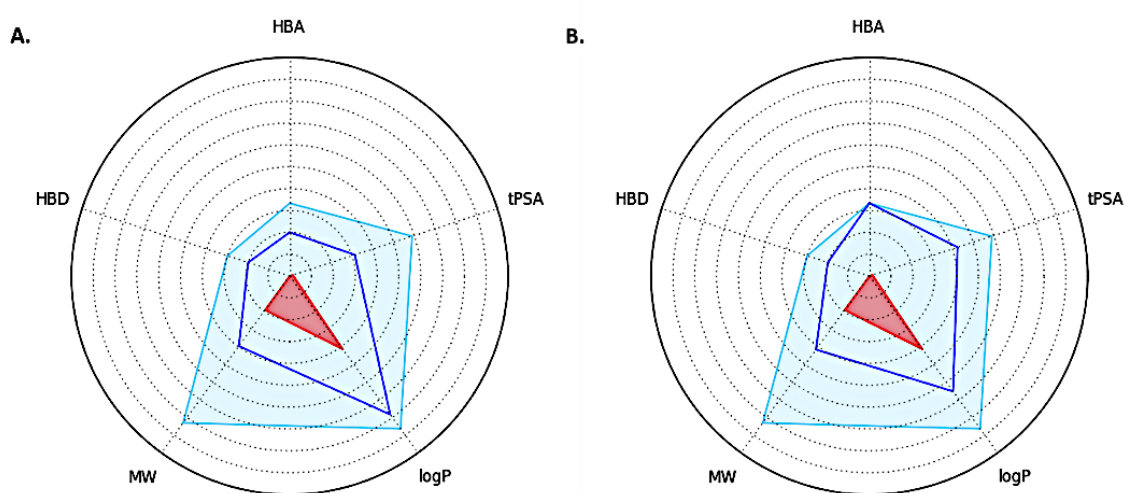


Figure 25. Compounds **18** and **26** (represented in A. and B. respectively), have passed the CNS filter which takes into consideration the assessment of their ability to pass the blood brain barrier, values of compounds **18** and **26** (blue line) fall within the CNS filter area (light blue)

3.4 Conclusions

Here, a successful computational strategy is reported for designing the first A₁R/A_{2A}R-PDE10A multi-target ligands with potential therapeutic utility for neurodegenerative diseases. A retrosynthetic approach was employed using MOE/RECAP, followed by target prediction and docking of the resulting library against the desired targets. 2-aminopyridine-3-carbonitriles have been identified as a series that showed agreement between both the ligand- and structure-based predictions of activity against A₁R, A_{2A}R and PDE10A. The synthesis of this series *via* a one-pot synthetic scheme was pursued experimentally. As a result, compounds **18**, **26**, **31**, and **35** were validated as A₁R/A_{2A}R-PDE10A multi-target ligands with IC₅₀ values of 2.4, 3.2,

10.0, and 5.1 μ M against PDE10A, and binding to A₁R with K_i values of 294 and 34 nM (**18** and **26** respectively), and to A_{2A}R with K_i values of 41, 95, and 55 nM (**26**, **31**, and **35** respectively).

Furthermore, selectivity profiling of the synthesized 4,6-substituted 2-amino-pyridin-3-carbonitriles against other subtypes of both protein families showed that the multi-target ligand **18** exhibited a minimum of 2-fold selectivity over all tested off-targets. In addition, compounds **18** and **26** exhibited the desired multi-target profile against A₁R, A_{2A}R and PDE10A, which would serve as good starting points for further functional efficacy assessment and analog modification for the improvement of selectivity. In particular, investigating the signal transduction profiles of these compounds using techniques previously described,[400] as well as evaluating functional effects in cAMP signaling assays may help determine if these compounds do indeed provide synergistic elevations in intracellular cAMP. The functional profile investigated here is likely to elevate cAMP levels synergistically *via* the combination effect on multiple targets simultaneously (as an A₁R antagonist/A_{2A}R agonist, and PDE10A inhibitor).

In summary, a computational approach for the design of multi-target ligands has been investigated. The method was validated experimentally *via* the synthesis and pharmacological evaluation of 2-aminopyridine-3-carbonitriles as A₁R/A_{2A}R-PDE10A ligands. This approach is generally applicable to a wide range of multi-target ligand design problems, across disease areas and target families.

4 Structure-based identification of dual ligands at A_{2A}R and PDE10A with anti-proliferative effects upon lung carcinoma cell-lines

4.1 Introduction

Targeting compounds that elevate cAMP concentrations *via* the agonism of A_{2A}R and the inhibition of PDE10A is a promising way to inhibit cancer cell proliferation.[98] This is achieved through the A_{2A}R-G α s-adenylate cyclase axis, while further promoting cAMP elevation by the inhibition of its breakdown *via* PDE10A. In fact, elevation of cAMP has been shown to exhibit anti-proliferative effects,[101–106] in various cell types. Given that A_{2A}R and PDE10A are both expressed in non-small cell lung cancer (NSCLC) cell-lines, [130, 131] then their targeting could be a promising anti-proliferative strategy in NSCLC. Indeed, there is the opportunity for multi-target approaches to improve therapy in NSCLC given that the first generation therapies have failed due to the use of single target agents that allow the survival of cancer cells *via* other pathways.[28]

In chapter 3, triazoloquinazolines were predicted as dual ligands at the A_{2A}R and PDE10A.[338] It has been challenging however, to identify compounds, which elevate cAMP at these targets.[402–404] In this chapter, this problem is addressed by docking triazoloquinazolines, which are known PDE10A inhibitors, into the orthosteric site of an active form of A_{2A}R.[319] Subsequently, MD simulation analysis was performed in order to identify a conformational descriptor that characterizes A_{2A} receptor activation.[320, 321, 405] Following this, the compounds are validated as A_{2A}R agonists in relevant biochemical assays and tested for their anti-proliferative effects in lung carcinoma cell-lines.

Cyclic adenosine monophosphate (cAMP) is a second messenger that has a major role in transduction and cell signaling in several pathways and biological systems.[96] cAMP elevation may be achieved *via* the activation of the adenylate cyclases by Gs proteins, and the inhibition of cAMP-degrading phosphodiesterases.[406] Given that cAMP intracellular signaling has been shown to inhibit proliferation, this signaling pathway would impact the survival and growth of cancer cells.[406] Indeed, compounds that elevate cAMP demonstrate therapeutic benefits in several cancer-related diseases such as breast cancer, colon cancer, lung cancer, glioblastoma etc.[407] In some cases, however, they exhibited pro-proliferative effects,

[99–101] whereas in many other cases anti-proliferative effects[101–106] depending on the cell type. For instance, cAMP elevation leads to the blockade of growth factor-stimulated cell growth *via* the inhibition of the mitogen-activated protein (MAP) kinase cascade, resulting in anti-proliferative effects in mesangial cells, fibroblasts, and smooth muscle cells. In contrast, cAMP elevation exhibits pro-proliferative effects in other cell types such as hepatocytes, thyroid cells, and PC12 cells.[99–106]

A_{2A}R is expressed in both lung adenocarcinoma and squamous carcinoma cell-lines, which are two histologically distinct types of non-small cell lung cancer (NSCLC cell-lines).[130, 131] Likewise, PDE10A is overexpressed in lung adenocarcinoma, and its inhibition was found to suppress growth,[132] demonstrating a correlation between the levels of overexpression and survival.[408] Also, a recent study suggests that the cAMP signaling pathway contributes to the suppression of cell growth in NSCLC.[409] Therefore there is the possibility of targeting both A_{2A}R and PDE10A as an anti-proliferative strategy for NSCLC, given that the agonism of the A_{2A}R and inhibition of PDE10A lead to cAMP elevation.

Indeed, dual PDE inhibition and A_{2A}R activation *via* compound combinations has exhibited synergy in cAMP elevation and been observed to inhibit proliferation of multiple myeloma and diffuse large B-cell lymphoma, as well as induce apoptosis.[98] Based on these results, the anti-proliferative effects of dual ligands at A_{2A}R and PDE10A in lung adenocarcinoma and squamous cell carcinomas cell-lines were explored as promising avenues for investigation.

Many virtual screening protocols have been reported in the literature, implementing either ligand- or structure-based approaches for the design of PDE10A inhibitors or A_{2A}R agonists. Examples of ligand-based protocols include *in silico* target prediction, pharmacophore-based and fragment-based approaches and comparative molecular field analysis (CoMFA).[386, 402, 403, 410–412] Docking, as a structure-based approach, has also been employed for the design of either PDE10A inhibitors or A_{2A}R agonists.[413] In addition, many molecular dynamics approaches have been used to investigate the conformational dynamics at the A_{2A} adenosine receptor or PDE10A.[350, 404, 413–418] None of the reported protocols, however rationalize or correctly predict the functional activity of ligands against the targets of interest, in particular A_{2A}R, which is addressed in this work.

Here, a novel structure-based approach for identifying ligands that activate the A_{2A}R while simultaneously inhibiting the PDE10A is devised. Given that PDE10A is an enzyme, compounds that target its active site would most likely confer inhibition. The real challenge

lies in finding compounds that activate the A_{2A}R given that binding to its orthosteric site may not guarantee the desired functional activity. For this reason, the design of A_{2A}R agonists was focused on the objective of finding dual-target ligands.

Triazoloquinazolines have been predicted as a chemical series of dual ligands with activity at A_{2A}R and PDE10A by ligand- and structure- based techniques (as described in chapter 3.)[338] So known PDE10A inhibitors, belonging to this chemical series, were docked into the orthosteric site of the active form of the A_{2A}R protein crystal structure (PDB ID: 2YDO). This crystal structure that enriches A_{2A}R agonists over antagonists and inactives was selected for docking studies. Experiments were performed by the Ladds group members at the pharmacology department in the University of Cambridge using both yeast-based[419] and mammalian cell-based assays[400] to demonstrate that triazoloquinazolines are indeed A_{2A}R agonists. A particular focus in the structure-based computational approach was on the key interacting residues, which are reported in the literature to discriminate between agonist and antagonist activity of A_{2A}R ligands. It is reported that[320, 405] Val₈₄, Leu₈₅, Leu₂₄₉, Ser₂₇₇ and His₂₇₈ show significant differences between the active and inactive conformations of the A_{2A}R protein crystal structures. In particular, Val₈₄ in TM3 (Transmembrane Helix 3), an amino acid residue that is located in the orthosteric site, has to move upon agonist binding owing to a steric clash with the agonist, which may contribute to the 2Å shift observed in H3 (Helix 3).[319–321] Hence, it is postulated that the motion of this residue upon A_{2A}R ligand binding might discriminate between agonist and antagonist activity, which has not previously been investigated by MD approaches.[412–417, 420]

Molecular dynamics (MD) analysis is used to determine the motion of Val₈₄ in TM3 and to deduce whether it is an essential requirement of A_{2A}R activation. In this work, the motion of Val₈₄ was found to be a good conformational descriptor for the characterization of receptor activation by the A_{2A}R ligands. Hence the MD-assisted approach provides a template for generating A_{2A}R agonists as part of the dual-target ligand design objective by identifying a shortlist of candidates that displace the Val₈₄ amino-acid residue.

Furthermore, the compounds showed concentration-dependent anti-proliferative effects in lung squamous cell carcinoma and lung adenocarcinoma cell-lines. These effects correlate with the co-expression of A_{2A}R and PDE10A and the increased cellular levels of cAMP, with the H1563 adenocarcinoma cell-line being the most sensitive given that it exhibits the highest combined expression of A_{2A}R and PDE10A. In particular, compound **36** shows a strong correlation

between the increase in potency for cAMP elevation and inhibition of proliferation in lung carcinoma cell-lines as the combined expression of A_{2A}R and PDE10A increases across the LK-2, H520, H1792 and H1563 cell-lines. The anti-proliferative effects that were exhibited by these dual-target compounds in a concentration-dependent manner suggest their potential clinical value.

4.2 Materials and Methods

4.2.1 Design approach of the dual ligands at the A_{2A}R and PDE10A

Triazoloquinazolines were shortlisted as candidates of dual ligands at the A_{2A}R and PDE10A, given that this chemical series displayed the highest frequency of prediction against the desired set of targets by the ligand- and structure- based techniques (see chapter 3).[338] However to elevate cAMP, the focus was to find dual ligands that are simultaneously agonists at A_{2A}R and inhibitors of PDE10A.

From the ZINC¹² database,[421] eleven purchasable triazoloquinazolines that were experimentally known PDE10A inhibitors, were found upon a search done for the triazoloquinazoline substructure with the following criteria: Uniprot ID: Q9Y233 and IC₅₀ < 10 µM. The triazoloquinazolines that were identified to meet these criteria had the following IDs: 3154141, 3141002, 6206233, 9937921, 9939949, 2968902, 14728559, 424907, 13229753, 44924158, and 8747709. These were downloaded for subsequent docking into the orthosteric site of the A_{2A}R protein crystal structure.

4.2.2 Selection of the A_{2A}R protein crystal structure for shortlisting triazoloquinazoline candidates as A_{2A}R agonists

All the active forms of the A_{2A}R protein crystal structure with the following PDB IDs (4UG2, 4UHR, 3QAK, 2YDO, and 2YDV)[319, 320, 405] and the inactive forms with the following PDB IDs (5IU4, 3UZA, 5K2A, 4EIY, 3EML, 5NM2, 5JTB, 5UVI, and 5UIG)[368, 398, 399, 422–426] were downloaded into MOE.[352] It has been reported in the literature that Val₈₄ in TM3, which is located in the orthosteric site, has to shift its position upon agonist binding owing to a steric clash with the ligand, which may contribute to the 2Å shift observed in H3.[319–321] To evaluate the change in the interaction upon agonist binding, the distance was calculated using MOE from a fixed amino acid residue to Val₈₄. The fixed amino acid residue selected was Leu₂₄₉ in TM6. This was achieved by aligning all the active and inactive forms of the A_{2A}R protein crystal structures (using the sequence editor > alignment> align/superimpose

option). Then, the mean RMSD value of all the aligned structures was calculated for Leu₂₄₉, which turned out to be low (0.40Å) confirming that it is reasonably static in its relative position. For each PDB ID of the active and inactive forms of the A_{2A}R crystal structures, the distance between the α -carbons of Val₈₄ in TM3 and Leu₂₄₉ in TM6 was measured in MOE using the measure> distances option. Table 6 lists all the Val₈₄-Leu₂₄₉ inter-residue distance values for the active and inactive forms of the A_{2A}R protein crystal structures. The inter-residue distances of the active forms ranged from 14.30 to 14.53Å, and for the inactive forms they ranged from 12.96 to 13.36Å. The largest displacement of the Val₈₄ residue was measured for the active form with PDB ID: 2YDO, and the distance was equal to 14.53Å, this is in comparison to the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4), which had the minimum distance (12.96Å). Given that Val₈₄ displayed the highest displacement from the Leu₂₄₉ residue in this crystal structure, it was hypothesized that this would allow an optimum conformational space for A_{2A}R agonists to explore when docked into its orthosteric site. Hence, the A_{2A}R protein crystal structure with the PDB ID: 2YDO was selected as the best candidate for shortlisting A_{2A}R agonists.

Table 6. The inter-residue distance values for Val₈₄-Leu₂₄₉ for all the active and inactive forms of the A_{2A}R protein crystal structures

Active forms of the A _{2A} R		Inactive forms of the A _{2A} R	
PDB ID	Val ₈₄ -Leu ₂₄₉ Distance (Å)	PDB ID	Val ₈₄ -Leu ₂₄₉ Distance (Å)
4UG2	14.30	5IU4	12.96
4UHR	14.38	3UZA	13.15
3QAK	14.46	5K2A	13.01
2YDO	14.53	4EIY	13.07
2YDV	14.35	3EML	13.36
		5NM2	13.06
		5JTB	12.98
		5UVI	13.07
		5UIG	12.98

4.2.3 Ligand Preparation

39 potent agonists and 38 potent antagonists of the A_{2A}R (Uniprot ID: P29274) with EC₅₀ and IC₅₀ values less than 1 μ M and confidence scores equal to 9 were manually extracted from ChEMBL. 133 A_{2A}R inactives were extracted from PubChem (using SQL query in Appendix D) and the eleven purchasable triazoloquinazolines were selected from the ZINC¹² database. The entire set of ligands were prepared for docking into the orthosteric site of the A_{2A}R protein crystal structure, with LigPrep 2.5[372] using the default settings and the Epik option, which introduces energy penalties associated with ionization and tautomerization.[335]

4.2.4 Receptor Preparation

Docking with Glide[134] was performed against the human A_{2A}R protein crystal structure (PDB IDs: 2YDO and 5IU4). The protein structures were prepared using the Protein Preparation Wizard of Maestro 9.3,[334] following the default protocol, which accounts for energy refinement, hydrogen addition, pKa assignment, and side-chain rotational isomer refinement. Resolved water molecules were discarded, and the structure was centered using the co-crystallized ligand as the center of the receptor grid generated for each protein structure. The co-crystal structures of A_{2A}R with Adenosine (PDB ID: 2YDO) and with ZM 241385 (PDB ID: 5IU4) were selected as target structures.

4.2.5 Enrichment of agonists by the A_{2A}R docking model (PDB ID: 2YDO)

In an attempt to validate the A_{2A}R docking model, the set of prepared A_{2A}R agonists, antagonists and inactives were docked using Glide against the prepared protein structure.

The Glide docking parameters used were extra precision (XP) and flexible ligand sampling, which obtained the best separation for the medians of docking score distribution for agonists versus antagonists and agonists versus inactives. This implies that this docking model enriches the agonists. Figure 26 shows the separation of the medians for the A_{2A}R docking model: A. -11.24 (agonists) B. -7.88 (antagonists) and C. -6.74 (inactives). Statistical analysis was performed with R[337] using a Mann-Whitney test on the agonist and antagonist docking score distributions, as well as agonist and inactive docking score distributions. The differences in medians were significant at a p value of less than 0.05 (script provided in [338]).

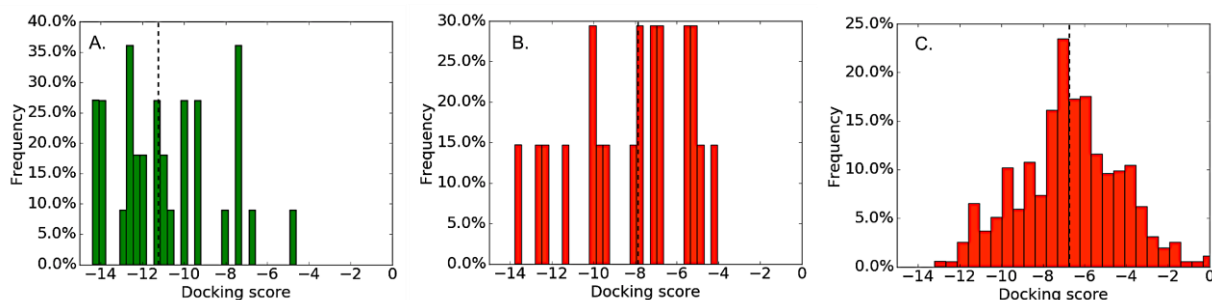


Figure 26. A good separation was obtained for the medians (dashed lines) of the docking score distributions for **A.** agonists (-11.24) **B.** versus antagonists (-7.88) and **C.** versus inactives (-6.74) confirming that the agonists are enriched. Statistical analysis was performed with R using a Mann-Whitney test[337] on the agonist and antagonist docking score distributions, as well as agonist and inactive docking score distributions. The differences in medians were significant for a p value less than 0.05.

4.2.6 Cut-off generation for compound selection as candidates of A_{2A}R agonists from the docking model

The Matthews correlation coefficient (MCC), which takes into account true and false positives (agonists) and negatives (antagonists), was computed (using a Python script[338]) for the docking scores of the agonists and antagonists against the A_{2A}R docking model. A search was performed for a docking score threshold that gave the highest MCC in order to shortlist promising candidates of A_{2A}R agonists, which displayed docking scores that are lower than the score with the highest MCC, and this gave a threshold of -7.33 for the A_{2A}R docking model.

4.2.7 Docking

The eleven purchasable triazoloquinazolines, which were prepared with LigPrep, were docked against the A_{2A}R protein crystal structure (PDB ID: 2YDO). The Glide docking parameters used were extra precision (XP) and flexible ligand sampling. The parameters were deduced from docking experiments using known actives and inactives against the protein-docking model. Six triazoloquinazolines (**36-41**) displayed docking scores that are lower than -7.33, which was the docking score with the highest MCC for the known agonists and antagonists. Their chemical structures are depicted in Figure 27. As a further step, compounds **36**, **39** and **40**, with the highest predicted affinities and the most potent agonists identified, compound **41**, which did not exhibit any agonist activity, CHEMBL3799351 (antagonist with an IC₅₀ = 4.35 nM and confidence score equal to 9), and CGS 21680 (the selective and potent A_{2A}R agonist), Adenosine (non-selective Adenosine receptor agonist), were docked into the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4) for MD simulation and analysis. The six

triazoloquinazolines (**36-41**) were then shortlisted for validation as A_{2A}R agonists in relevant biochemical assays.

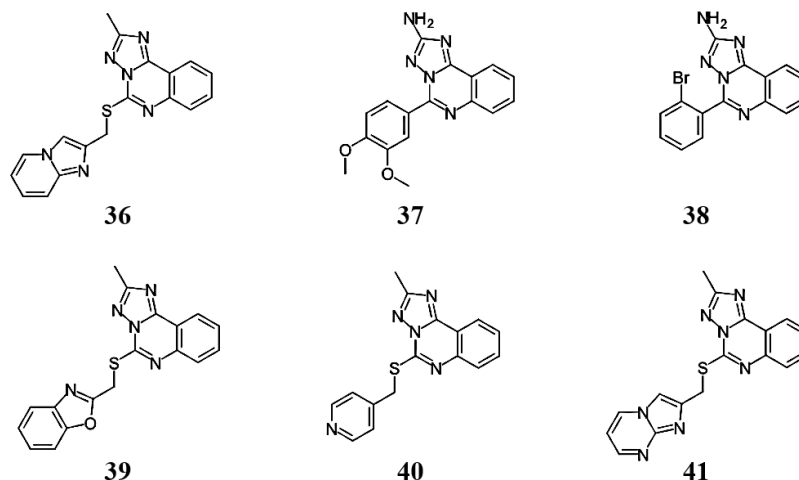


Figure 27. The chemical structures of the six triazoloquinazolines (**36-41**) that displayed docking scores that are lower than -7.33, which was the docking score with the highest MCC for the known agonists and antagonists. They were then then shortlisted for validation as A_{2A}R agonists in relevant biochemical assays.

4.2.8 MD Simulations

Based on a structural analysis of the available A_{2A}R crystal structures, the distance between the α -carbons of Val₈₄ in TM3 and Leu₂₄₉ in TM6 was selected to be investigated as a conformational descriptor for receptor activation. The two A_{2A}R co-crystallized structures of (PDB IDs: 5IU4 and 2YDO), which exhibited the largest difference in α -carbon distances between Val₈₄ in TM3 and Leu₂₄₉ in TM6 (12.96Å versus 14.53Å respectively), were selected for molecular dynamics simulation analysis. Subsequently, compounds **36**, **39**, **40**, and **41** that were docked into the orthosteric site of the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4) were subjected to a 100 ns MD simulation protocol. Likewise, CHEMBL3799351, CGS 21680, Adenosine, which were all docked into the orthosteric site of the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4), and the apo structure (PDB ID: 5UI4), were chosen for the same analysis. The starting structures were prepared using Maestro 9.3 following the default procedure for protein preparation. The protocol uses the "Cap termini" option, which adds the coordinates to the residue to have them in order. Next, "Analyze network" in the Interactive H-bond optimizer was used to check on the assignments of hydrogen orientations in the hydrogen-bonding network. They were subsequently optimized. All MD simulations described in this study were performed using Desmond 3.2 available in the Schrödinger software package.[265] An orthorhombic box was used to build the model

systems with periodic boundary conditions in an isothermal–isobaric ensemble with a constant number of particles (NPT ensemble). The system temperature was kept at 300 K, and the pressure was kept at atmospheric pressure. The definition of transmembrane regions was taken from the OPM database.[427] The receptor structures were embedded in a pre-equilibrated palmitoyl-oleoyl-phosphatidylcholine membrane (bilayer) and solvated with simple point charge water and 0.15 M NaCl. All other parameters were set on default values. The 100 ns simulations were carried out with Desmond 3.2 *via* command line on the computer cluster CALCULON (University of Cambridge) by using 20 central processing units. The trajectories obtained were analyzed with the software VMD,[428] and then plots were obtained for the RMSD values of His₂₅₀ in TM6, and the α -carbons distances between Val₈₄ in TM3 and Leu₂₄₉ in TM6 for the simulated systems over 100 ns using seaborn library[429].

4.2.9 Experimental validation of the virtual screening protocol

Ian Winfield (IW), Dewi Safitri (DS), and Sabrina Carvalho (SC) in the Ladds group, at the Pharmacology Department of the University of Cambridge, performed the experimental validation of triazoloquinazolines **36-41** as dual ligands at A_{2A}R and PDE10A, and tested the anti-proliferative effects of the compounds in lung carcinoma cell lines.

4.2.9.1 Materials

Triazoloquinazolines **36-41** were supplied from Ambinter (Orléans, France), and CGS 21680, NECA and ZM 241385 from Tocris Biosciences (Abingdon, UK). All compounds were stored in 10 mM stock solutions in DMSO. Rolipram was purchased from Cayman chemicals (Michigan USA), and other laboratory reagents were from Sigma-Aldrich (Poole, UK), of analytical grade.

4.2.9.2 Mammalian cell culture

CHO-K1 (gifted by Dr. Ewan St. John Smith, University of Cambridge, UK) CHO-K1-A_{2A}R and CHO-K1-A₃R cells (gifted by Prof. Karl-Norbert Klotz, University of Wuerzburg, Germany), were routinely cultured in Hams F-12 nutrient mix, supplemented with 10% fetal bovine serum (FBS). H520, H1563, H1792 and LK-2 cells (gifted by Dr. Whalid Khaled, University of Cambridge, UK) were grown in RPMI media + 10% FBS. All media was further supplemented with 1X antibiotic, antimycotic solution (Sigma Aldrich, Poole, UK). Culturing of all cell types was done at 37°C in a humidified atmosphere containing 5% CO₂. (work done by IW and DS)

4.2.9.3 Generation of CHO-K1 cell line stably expressing the A_{2A}R

CHO-K1 cells stably expressing the A_{2A}R cells were generated *via* transfection with 500 ng pcDNA3.1-A_{2A}R (cDNA.org), per well of a 24-well plate, which was performed with FuGENE HD (Promega, Wisconsin, USA), at a 1:3 (w/v) DNA:FuGENE ratio. Prior to adding 800 µg/ml G418 (Sigma Aldrich, Poole, UK), the cells were further cultured for 48 hours. Then every 48 hours, G418 containing media were replaced until foci of cells were attained, which were left to grow to 100% confluency. Afterwards, each well was tested for the ability of CGS 21680 to elevate cAMP, performing further culturing with appropriately responding clones as described (work done by IW).

4.2.9.4 Phosphodiesterase 10A inhibition assays

A PDE10A assay kit (BPS Bioscience, San Diego, CA) was used by IW to test the PDE10A inhibition of compounds **36-41** as described in the manufactures protocol. 400 pg of purified PDE10A was used per reaction, and the plates were read using a TECAN infinite M200.

4.2.9.5 Yeast methods

Generation of yeast strains was done by IW and SC according to previously reported protocols, and they have been routinely grown as previously described by Knight *et al.*[400] Yeast cells expressing either the A₁R, A_{2A}R, or A_{2B}R were treated with either NECA, CGS 21680 or compounds **36-41**, in order to measure the activity of each, as previously described.[400]

4.2.9.6 cAMP accumulation assays

Prior to assay, harvesting of cells was performed with trypsin containing 0.05% EDTA, they were then washed with PBS, and subsequently resuspended in stimulation buffer (PBS Proliferation assays containing 0.1% BSA and 25 µM rolipram). Seeding of cells was done at 2000 cells well⁻¹ of a 384-well white optiplate, and then they were stimulated at room temperature with compounds **36-41** (ranging 10 mM – 100 pM) for 30 minutes. The cells were subsequently lysed, and the measurement of cAMP levels was done using a LANCE cAMP detection kit (PerkinElmer), and the plates were read with a Mithras LB940 microplate reader (work done by IW).

4.2.9.7 Proliferation assays

To test the effect of compounds **36-41** upon proliferation, various cell types were seeded onto clear 96-well plates at proper densities for each; CHO-K1 (2000 cells well⁻¹), CHO-K1-A_{2A}R

(2000 cells well⁻¹), H520 (2500 cells well⁻¹), H1563 (2500 cells well⁻¹), H1792 (2500 cells well⁻¹), LK-2 (2500 cells well⁻¹). This was done in suitable media, and they were cultured for 24 hours. After the subsequent addition of compounds **36-41** (ranging 316 nM - 100 μ M), cells were allowed to grow further for 72 hours. Quantification of changes in cell number was done by adding 5 μ l CCK-8 reagent to each well, accompanied by incubation at 37°C for 1-3 hours. The determination of OD₄₅₀ was done using a Mithras LB940 micro-plate reader at 450 nm (work done by IW and DS).

4.2.9.8 RT-PCR

Extraction of RNA from H520, H1792, H1563 and LK-2 cells was done using a RNeasy[®]-4PCR Total RNA Isolation Kit (Life Technologies, Paisley, UK) by DS as per the manufacturer's instructions. Then, DNase I treatment was performed to remove the contamination by genomic DNA. Subsequently, the quantification of the degree of purity of RNA samples was performed using a NanoDrop[™] Lite spectrophotometer (Thermo Scientific, UK). The samples that were used in cDNA synthesis are those of yields >100 ng/ μ L and A_{260/280} ratios >1.9. The cDNA synthesis was done using a QuantiTect reverse transcription kit (Qiagen, Manchester, UK), for which a total of 1 μ g of freshly isolated RNA was consumed per reaction. RT-PCR was subsequently implemented according to what has been previously reported[430]. The RT-PCR that has been done used gene specific primers to human: *GAPDH* (Sense 5'-TGCACCACCAACTGCTTAGC-3'; Antisense 5'-GGCATGGACTGTGGTCATGAG-3'), *A₁R* (Sense 5'-CCACAGACCTACTTCCACACC-3'; Antisense 5'-TACCGGAGAGGGATCTTGACC-3', Primerbank ID - 115305570C1[431]), *A_{2A}R* (Sense 5'-CGCTCCGGTACAATGGCTT-3'; Antisense 5'-TTGTTCCAACCTAGCATGGGA-3', Primerbank ID - 156142194C1[431]), *A_{2B}R* (Sense 5'-TGCACTGACTTCTACGGCTG-3'; Antisense 5'-GGTCCCCGTGACCAAACCTT-3', Primerbank ID - 22907046C1[431]), *A₃R* (Sense 5'-GGCCAATGTTACCTACATCACC-3'; Antisense 5'-CCAGGGCTAGAGAGACAATGAA-3', Primerbank ID - 4501953A1[431]) and *PDE10A* (Sense 5'-TGA TGACTTTTCTCTCGACGTTG-3'; Antisense 5'-AAGCCACCTACACAGTGTCTC-3', Primerbank ID - 359465520C1[431]). Then, gel electrophoresis (using 2% agarose gels) was performed to resolve PCR products. The imaging of gels was subsequently done using a G Box iChemi gel documentation system employing GeneTools analysis software (Syngene, Cambridge, UK) and densitometry.

4.2.9.9 Data analysis

Data analysis was performed by the Ladds group members using GraphPad Prism 6e (San Diego, CA). All data for β -galactosidase assays were normalized to the responses resulting from NECA stimulation, whereas the data for cAMP inhibition/accumulation assays were normalised to those obtained upon stimulation with 100 μ M Forskolin or CGS 21680. As for proliferation assays, the normalization of all data was done relative to the responses obtained upon treating cells with 1% (v/v) DMSO. Subsequently, a three-parameter logistic equation was used for fitting each set of normalized data β -galactosidase or cAMP data, in order to calculate pEC₅₀/pIC₅₀ and E_{max} values. Also the fitting of the proliferation data was done using a three-parameter logistic equation constraining the basal value to 100 and the system maximum to the I_{Max} value obtained for compound **37**, since it elicited the maximum inhibition of cellular proliferation in all cell types tested. A one-way ANOVA with Dunnett's post-test, or Student's t-test was used to assess the statistical significance for all assays, where $p < 0.05$ was considered to be significant.

4.3 Results and Discussion

4.3.1 Method of selecting triazoloquinazolines as candidates of dual ligands at A_{2A}R and PDE10A

Triazoloquinazolines were identified as a compound series that showed the highest frequency of prediction at A_{2A}R and PDE10A by the ligand- and structure- based techniques.[338] For the purpose of finding dual-target ligands that elevate cAMP, the focus was on ligands that simultaneously activate the A_{2A}R and inhibit the PDE10A. In particular, additional attention was given to identifying A_{2A}R agonists.

From the ZINC¹² database,[421] six purchasable triazoloquinazolines (see methods for details), which are known to inhibit the PDE10A with IC₅₀ values ranging from 15 nM to 10 μ M, were shortlisted. This was done by docking them into the orthosteric site of the A_{2A}R protein crystal structure (PDB ID: 2YDO), which displayed the highest displacement of the Val₈₄ residue, (Table 6) an amino acid residue whose motion is essential for A_{2A}R activation. The Val₈₄-Leu₂₄₉ inter-residue distances were found for the active forms of the A_{2A}R protein crystal structure reported in PDB, and compared to the inactive forms in PDB. Hence, the docking model (PDB ID: 2YDO) was selected based on this criterion, and it was able to enrich A_{2A}R agonists over A_{2A}R antagonists and inactives (refer to methods for the choice of this protein

structure and the validation of the docking model used). This is in line with a previous study done by Rodríguez *et al.*,[404] where the A_{2A}R crystal structure (PDB ID: 2YDO) displayed the highest enrichment factor value (EF1%) for docked agonists over the other active crystal structures of A_{2A}R. The 2YDO crystal structure enriched agonists 63.5-fold better than random and 2.9-fold better than antagonists (63.5% versus 21.9%). However, this docking approach failed to find any A_{2A}R agonists (which used three active structures PDB IDs: 2YDO, 2YDV, and 3QAK). The authors rationalized this as the result of bias of the chemical libraries towards A_{2A}R antagonists over agonists.

In this approach, the selection of the six triazoloquinazolines, as promising candidates of A_{2A}R agonists was based on the docking scores that are lower than the score with the highest MCC for the A_{2A}R docking model, which was -7.33 (see methods for details). Compound **39** exhibited the most favorable docking score (highest predicted affinity) among the six shortlisted compounds. Compounds **36-41** were screened against PAINs (PAN Assay Interference Compounds)[339] using FAFDrug3,[313] and none of the compounds exhibited any potential PAINs liability.

4.3.2 Analysis of the predicted binding modes of representative triazoloquinazolines 36-39 shortlisted for experimental validation

Docking studies predicted molecular interactions characteristic of triazoloquinazolines with the A_{2A}R protein crystal structure (PDB ID: 2YDO). Compounds **36-38**, were predicted to be selective A_{2A}R ligands, which could be attributed to their interactions with His₂₅₀ (Figure 28).[432, 433] This residue is located in the core region of the receptor and part of a sub-pocket formed by Leu₈₅, Met₁₇₇, Trp₂₄₆ and Leu₂₄₉. Despite the fact that it is conserved among the A₁R and the A_{2A}R subtypes (as suggested by a recent study, due to the high conservation of amino acid residues in the adenosine receptor subtypes), the subtype selectivity might not be attributed to the receptor-specific amino acid residues, but rather conformational differences.[363] In addition, given that mutation experiments have failed so far to highlight any receptor-specific amino acid residues responsible for subtype selectivity, this would further support the hypothesis.[433, 434] Hence, the selectivity of A_{2A}R agonists could be attributed to the conformational change of the His₂₅₀ amino acid residue that contributes to shaping the orthosteric site to favor their selectivity.[363] Indeed, the interaction with this residue is only observed for the selective A_{2A}R co-crystallized agonists, CGS 21680 (PDB ID: 4UHR) and UK432097 (PDB ID: 3QAK) but not the non-selective co-crystallized agonists NECA (PDB

ID: 2YDV) and adenosine (PDB ID: 2YDO). It has been reported, however, that the occurrence of this interaction cannot discriminate between agonist and antagonist activity.[433, 434] Hence, the predicted interactions with His₂₅₀ might serve to quantify the binding in the lipophilic sub-pocket formed by these residues, suggested to be the driving force for A_{2A}R subtype selectivity.

As shown in Figure 28, compound **36** forms H-bonds *via* the nitrogen of the quinazoline ring with the Asn₂₅₃, and *via* the imidazo ring with Glu₁₆₉. Its triazole ring is π -stacked with Phe₁₆₈, and its phenyl group in the quinazoline moiety is π -stacked with His₂₅₀. The overlaid compounds **37** and **38**, H-bond *via* their amino groups with Glu₁₆₉ and Asn₂₅₃, and their triazole rings are π -stacked with Phe₁₆₈, whereas their overlaid 3,4-dimethoxy-phenyl and 2-Bromophenyl (for **37** and **38** respectively), are π -stacked with His₂₅₀. As for compound **39**, it exhibits H-bonding *via* the nitrogen of the triazole ring with Asn₂₅₃. The triazole ring is π -stacked with Phe₁₆₈, and the phenyl group of the 1,3-benzoxazole is π -stacked with Tyr₂₇₁.

The molecular interactions predicted for the selective A_{2A}R agonists **36–38** are consistent with the observed interactions between the co-crystallized ligands and the active A_{2A}R crystal structures (PDB IDs: 4UG2, 4UHR,[320] 3QAK,[405] 2YDO and 2YDV[319]). As for compound **39**, its extra interaction with Tyr₂₇₁ is similar to that exhibited by UK432097 but not to any of the reported co-crystallized antagonists.[120–125, 127, 128] The compounds were not predicted to display all the interactions exhibited by the agonist co-crystallized ligands[319, 320, 405] and in particular the Thr₈₈ and Ser₂₇₇ interactions, which are also characteristic of the ZM 241385 antagonist.[418] Hence, these interaction types are not characteristic of agonist activity. However, it has been reported in the literature that mutating these residues has a stronger influence on agonist activity than upon the antagonist activity of the A_{2A}R ligands, but not on the binding to the A_{2A}R.[120] As for the co-crystallised A_{2A}R antagonists (PDB ID: 5IU4,[399] 3UZA,[422] 5K2A,[423] 4EIY,[368] 3EML,[398] 5NM2,[424] 5JTB,[425] 5UVI, and 5UIG)[426] they show only interactions with Phe₁₆₈, Asn₂₅₃, and Glu₁₆₉ residues. Therefore, the type of predicted interaction is not indicative of receptor activation by the triazoloquinazolines. However, the docking model used enriched A_{2A}R agonists over A_{2A}R antagonists and inactives. This suggested an investigation (using MD) whether the His₂₅₀ movement would differ between selective versus non-selective A_{2A}R agonist binding and also to investigate whether the motion of Val₈₄ would vary upon agonist and antagonist binding. This would allow discrimination between these different classes of compounds.

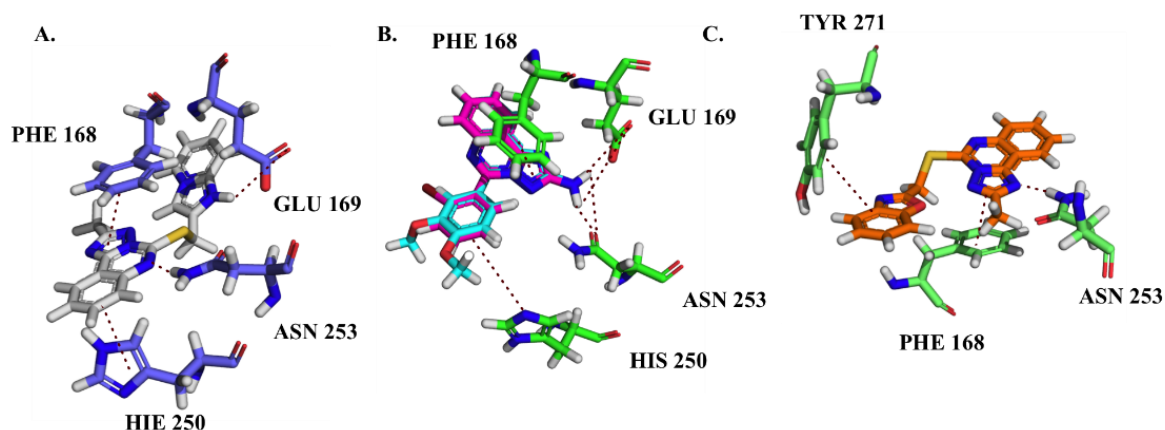


Figure 28. Docking studies predicted molecular interactions characteristic of triazoloquinazolines with the A_{2A}R protein crystal structure (PDB ID: 2YDO): **A.** compound **36** H-bonds *via* the nitrogen of the quinazoline ring with Asn₂₅₃, and *via* the imidazo ring with Glu₁₆₉, and the triazole ring is π -stacked with Phe₁₆₈, and the phenyl group in quinazoline is π -stacked with His₂₅₀ **B.** the overlaid compounds **37** and **38** H-bond *via* their amino groups with Glu₁₆₉ and Asn₂₅₃, and the triazole rings are π -stacked with Phe₁₆₈, the 3,4-dimethoxy-phenyl and the 2-Bromophenyl substituents of compound **37** and **38** are π -stacked with His₂₅₀ **C.** compound **39** H-bonds *via* the nitrogen of the triazole ring with Asn₂₅₃ and the triazole ring is π -stacked with Phe₁₆₈, and the phenyl group of the 1,3-benzoxazole is π -stacked with Tyr₂₇₁. The molecular interactions predicted for the selective A_{2A}R agonists **36–38** are consistent with observed interactions between co-crystallized ligands with the active form of the A_{2A}R protein crystal structures (PDB IDs: 4UG2, 4UHR, 3QAK, 2YDO and 2YDV).

4.3.3 MD simulation suggests that the conformational change of the His₂₅₀ residue contributes to shaping the orthosteric site pocket to favor selectivity for A_{2A}R agonists

In order to investigate further whether the conformational change of His₂₅₀ contributes to shaping the orthosteric site to favor the selectivity for agonists at A_{2A}R, an RMSD analysis was performed on this residue over a 100 ns MD simulation. This is a commonly used approach to determine conformational flexibility upon ligand binding.[435, 436] The analysis was performed for compounds **36** and CGS 21680 (the selective and potent A_{2A}R agonists) and compound **39** and Adenosine (the non-selective adenosine receptor agonists) docked to the A_{2A}R structure (PDB ID: 5IU4). The RMSD values of His₂₅₀ over the 100 ns simulation time are represented as the moving average trend-lines (bin-size of 10 frames) in Figure 29. Similar trend is seen in the case of compounds **36** and CGS 21680, wherein both RMSD values of His₂₅₀ increase over the simulation time converging towards similar RMSD values. This indicates that a conformational change is occurring in this residue upon selective A_{2A}R agonist binding. The greater increase in RMSD values observed for compound **36** in comparison to CGS 21680 could be related to the higher degree of selectivity that this compound exhibits as

an A_{2A}R agonist over the other adenosine receptor subtypes. This is in contrast to CGS 21680 that exhibits agonist activity against the A₃R receptor subtype (refer to section 4.3.6). This is not the case for compounds **39** and Adenosine (non-selective agonists) where the RMSD values are relatively constant over the 100 ns simulation. The RMSD distributions for the His₂₅₀ residue in the last 50 ns of the MD simulation performed for compounds **36**, **39**, CGS 21680, and Adenosine docked to the A_{2A}R, are plotted in Figure 30. The separation in RMSD value distributions for the selective A_{2A}R agonists (**36** and CGS 21680) versus the non-selective A_{2A}R agonists (**39** and Adenosine) is clearly illustrated. A Kolmogorov-Smirnov test, which included statistical analysis of the pair-wise RMSD distributions of His₂₅₀ in selective versus non-selective agonist bound A_{2A}R structures was performed. This yielded a p value less than 2.2×10^{-16} , indicating that the difference between the two types of distributions is statistically significant. Furthermore, the RMSD values of His₂₅₀ for the A_{2A}R structures bound to compound **36** and CGS 21680 were compared to a reference residue (Leu₂₄₉), over the 100 ns MD simulation, as illustrated in Figure 31. It is evident that the RMSD values of Leu₂₄₉ are relatively constant for both A_{2A}R structures bound to compound **36** and CGS 21680 over the 100 ns simulation. This emphasizes that the increase in RMSD values for His₂₅₀ is characteristic of its conformational change upon selective A_{2A}R agonist binding. Hence, the comparison in RMSD values for the His₂₅₀ residue in A_{2A}R structures bound to selective versus non-selective A_{2A}R agonists suggests that the conformational change of this residue contributes to shaping the orthosteric site pocket to favor the selectivity of A_{2A}R agonists.

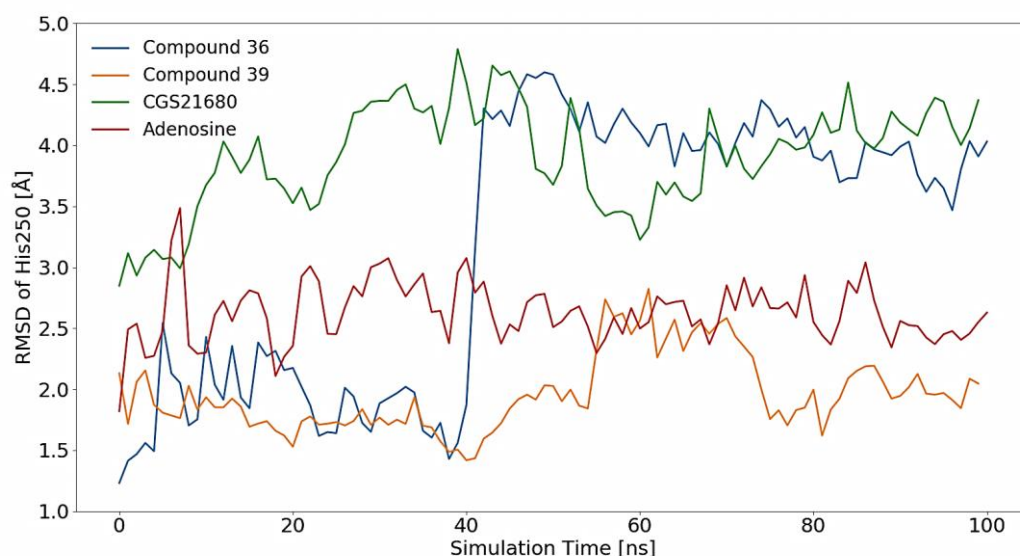


Figure 29. The moving average trend-lines (bin-size of 10 frames) of RMSD values of His₂₅₀ for compounds **36**, **39**, and CGS 21680 (the selective and potent A_{2A}R agonist) and Adenosine (the non-selective adenosine receptor agonist), docked to the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4) over a 100 ns simulation. The RMSD values of His₂₅₀ for the

docked structures of compounds **36** and CGS 21680 (the two selective A_{2A}R agonists) behave similarly by increasing and converging towards similar RMSD values. This indicates that a conformational change occurs in this residue upon A_{2A}R agonist binding. In contrast, the RMSD values in the case of the docked compounds **39** and Adenosine (the non-selective adenosine receptor agonists) are relatively constant. Hence, this suggests that the conformational change of this residue contributes to shaping the orthosteric site pocket to favor the selectivity of A_{2A}R agonists.

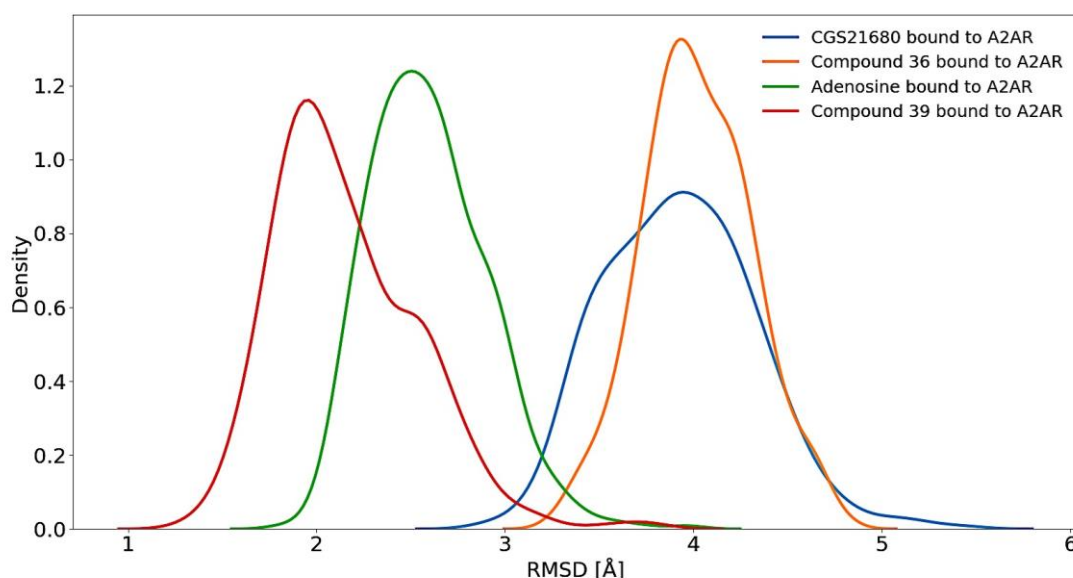


Figure 30. The RMSD distributions for the His₂₅₀ residue in the last 50 ns of the MD simulation performed for compounds **36**, **39**, CGS 21680, and Adenosine, docked to the A_{2A}R. The separation in RMSD value distributions for the selective A_{2A}R agonists (**36** and CGS 21680) versus the non-selective A_{2A}R agonists (**39** and Adenosine) is statistically significant by Kolmogorov-Smirnov test, for the pair-wise RMSD distributions (selective versus non-selective agonist bound A_{2A}R structures) with a p value less than 2.2×10^{-16} .

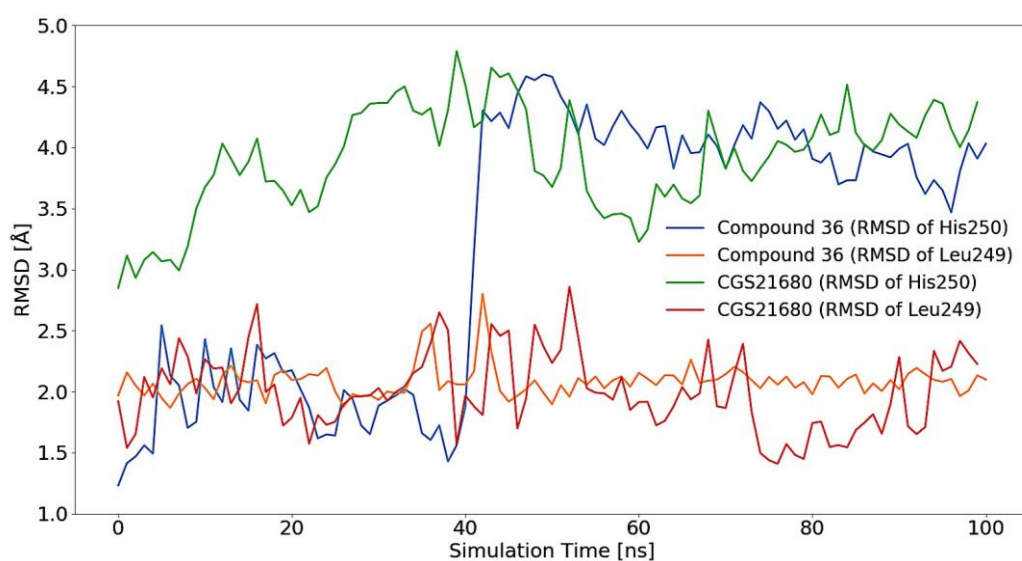


Figure 31. The moving average trend-lines (bin-size of 10 frames) of RMSD values of His₂₅₀ for the A_{2A}R structures bound to compound **36** and CGS 21680 were compared to those of a

reference residue (Leu₂₄₉) over a 100 ns MD simulation. The RMSD values of Leu₂₄₉ are relatively constant for both A_{2A}R structures bound to compound **36** and CGS 21680 emphasizing that increase in RMSD values for His₂₅₀ is characteristic of its conformational change upon selective A_{2A}R agonist binding.

4.3.4 MD Simulation analysis reveals the shift in Val₈₄ as a requirement for receptor activation by A_{2A}R ligands

The analysis of the active and inactive forms of the available A_{2A}R crystal structures is in accordance with reports in the literature, which mention that Val₈₄ in TM3 has to shift by approximately 2Å upon agonist binding to avoid a steric clash between the ligand and the receptor.[319–321] Hence, this gave rise to the hypothesis that the motion of this residue might discriminate between agonist and antagonist binding. Therefore, a 100 ns MD simulation was performed for each of the co-crystallized structures (PDB IDs: 5IU4 and 2YDO), which exhibited the largest differences in distance between the α -carbons of Val₈₄ in TM3, and Leu₂₄₉, a relatively fixed residue in TM6 (12.96Å versus 14.53Å respectively, see methods for details). The same analysis was carried out for the apo structure of the A_{2A}R (PDB ID: 5IU4), the docked triazoloquinazolines **36**, **39**, and **40** with the highest predicted affinities (and the most potent agonists identified by assays performed in the CHO-K1-A_{2A}R cells), compound **41** (which did not exhibit any agonist activity), ChEMBL3799351 (a potent antagonist), and CGS 21680 (the selective and potent A_{2A}R agonist). All of these compounds were docked into the orthosteric site of the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4).

Figure 32 shows the moving average trend-line (bin-size of 20 frames) for the distances between the α -carbons of Val₈₄ in TM3 and Leu₂₄₉ in TM6, for each trajectory of the apo structure and all the docked and co-crystallized structures. Interestingly, in the last 50 ns of the simulation, the antagonist-bound A_{2A}R system maintained the distance observed for the inactive crystal structure, whereas the agonist-bound systems resembled the properties of the active receptor conformation. As for the apo structure, this exhibited an intermediate distance between the agonist and antagonist bound structures. The docked structures, CGS 21680, compounds **36**, **39**, and **40** behaved similarly in particular over the last 50 ns (when considering the average distance between α -carbons of Val₈₄ and Leu₂₄₉) by displaying an increase in their average distance values ($14.04 \pm 0.51\text{\AA}$, $14.08 \pm 0.37\text{\AA}$, $14.61 \pm 0.67\text{\AA}$, and $14.84 \pm 0.59\text{\AA}$ respectively) in comparison to the apo structure simulation ($12.93 \pm 0.44\text{\AA}$). The values converged towards the average distance value ($14.84 \pm 0.34\text{\AA}$) of the active protein crystal structure (PDB ID: 2YDO). In contrast, the potent antagonist ChEMBL3799351, which was

docked to the A_{2A}R protein crystal structure (PDB ID: 5UI4), showed a slight decrease in its average distance values in comparison to the simulated apo structure. It exhibited similar average distance value ($11.45 \pm 0.60 \text{ \AA}$) to that of the co-crystallized ZM 241385 antagonist (PDB ID: 5UI4) ($11.44 \pm 0.51 \text{ \AA}$) in the last 50 ns. As for compound **41** and the apo-structure, their average distance values were alike over the last 50 ns ($12.97 \pm 0.54 \text{ \AA}$ and $12.93 \pm 0.44 \text{ \AA}$ respectively). This could serve as an explanation of why compound **41** did not exhibit any agonist activity.

Violin plots were obtained (Figure 33) for the α -carbon distances between Val₈₄ and Leu₂₄₉ for the remaining 50 ns of the simulation for representative agonist/antagonist systems (agonist bound to 2YDO, compound **39**, CGS 21680, antagonist (ZM 241385), other potent antagonist (CHEMBL3799351)). These plots display the density distributions for the measured distances, and highlight the larger inter-residue distance characterizing the agonist binding as compared to antagonist binding. A statistical analysis was performed on the last 50 ns distance distributions, using a Mann-Whitney test and a Kolmogorov-Smirnov test for the representative agonist/antagonist pairs displayed in Figure 33. The differences in medians of the distance distributions for each of the agonist/antagonist pairs were significant at a p value less than 0.05, and the p value for the Kolmogorov-Smirnov test was less than 2.2×10^{-16} . Hence, the increase in the distance between Val₈₄ and Leu₂₄₉ residues caused upon A_{2A}R agonist binding serves as a useful conformational descriptor for receptor activation by the A_{2A}R ligands.

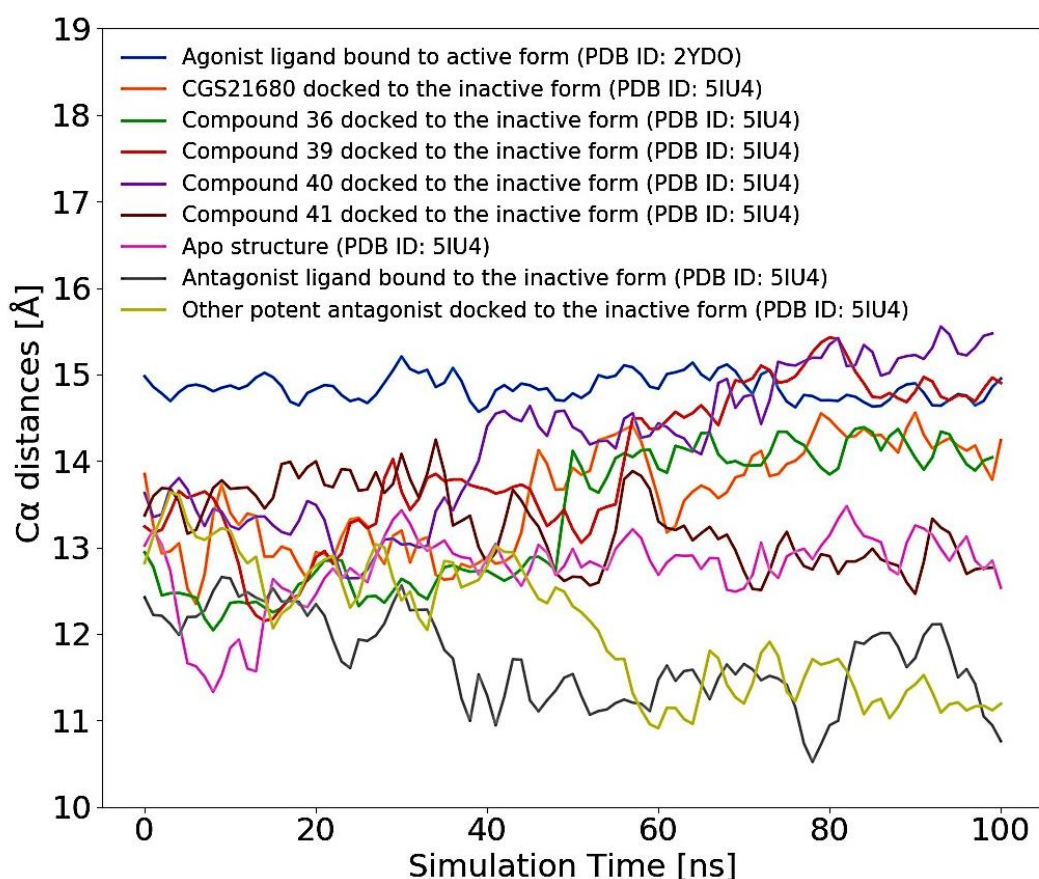


Figure 32. The moving average trend-lines (bin-size of 20 frames) are for the Val₈₄-Leu₂₄₉ C α distances of the apo structure (PDB ID: 5IU4) and the docked and the co-crystallized structures (PDB ID: 5IU4 and 2YDO) of the A_{2A}R over a 100 ns simulation. Compounds **36**, **39**, **40**, and **41**, a potent antagonist (CHEMBL3799351), and the selective and potent A_{2A}R agonist (CGS 21680) are docked into the inactive form of the A_{2A}R protein crystal structure (PDB ID: 5IU4). The variation in computed distances for compounds **36**, **39**, **40** and CGS 21680 were similar, where all increased in their average distances time moving towards the average distance observed in the dynamics of the active protein crystal structure (PDB ID: 2YDO). The average distances for compound **41**, and the apo-structure were similar, which explains why compound **41** did not exhibit any agonist activity. Hence, the increase in the Val₈₄-Leu₂₄₉ inter-residue distance upon A_{2A}R agonist binding serves as a good conformational descriptor for receptor activation by the A_{2A}R ligands.

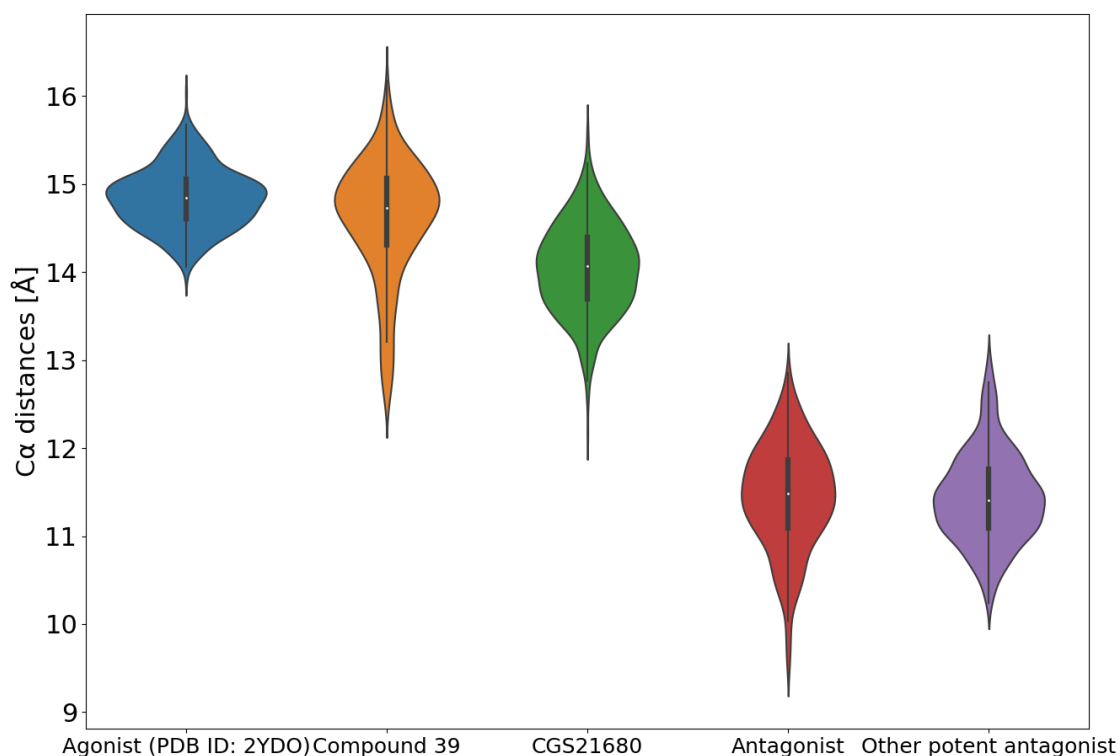


Figure 33. Violin plots for the α -carbon distances between Val₈₄ and Leu₂₄₉ for the last 50 ns of the MD simulations. It can be seen that larger inter-residue distances characterize agonist binding as compared to antagonist binding. Hence, the increase in the Val₈₄-Leu₂₄₉ inter-residue distance upon docking of the A_{2A}R agonists to the inactive form of the A_{2A}R protein crystal structure serves as a useful conformational descriptor for receptor activation by the A_{2A}R ligands.

4.3.5 Confirmation of triazoloquinazolines as PDE10A inhibitors

The PDE10A inhibitory activity was experimentally confirmed for the triazoloquinazolines **36-41** and the trend in potencies was comparable to what has been previously suggested by Kehler *et al.*[382] The rank of potencies for the six triazolquinazolines was **36** > **41** = **39** > **40** > **38** = **37** (Table 7 and Figure 34). The activity of CGS 21680 was tested and it was shown that it did not exhibit any PDE10A inhibitory activity.

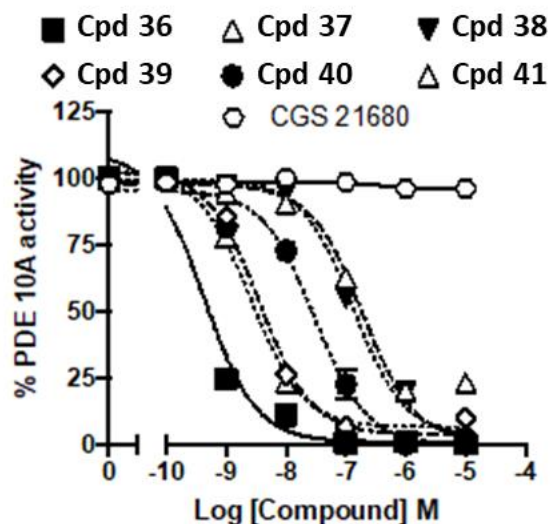


Figure 34. Dose response curves for PDE10A inhibition by triazoloquinazolines **36-41** and CGS 21680. Data are represented as the mean of 6 individual replicates \pm SEM (standard error of the mean). The rank of potencies obtained for the tested triazoloquinazolines was **36** > **41** = **39** > **40** > **38** = **37**. CGS 21680 did not inhibit PDE10A (plot generated by the Ladds group members).

Table 7. pIC_{50} values for PDE10A inhibition by the triazoloquinazolines **36-41** identified as dual ligands at PDE10A and $\text{A}_{2\text{A}}\text{R}$

Compound	pIC_{50}
36	9.87 ± 0.10
37	6.71 ± 0.03
38	7.01 ± 0.07
39	8.47 ± 0.04
40	7.49 ± 0.10
41	8.62 ± 0.12

4.3.6 Experimental validation of triazoloquinazolines as A_{2A}R agonists

Yeast strains, which have been previously characterized, were used in order to validate triazoloquinazolines **36-41** as A_{2A}R agonists and assess their selectivity against the human A₁R, A_{2A}R and A_{2B}R subtypes.[400] Stimulating each strain with the non-selective adenosine receptor agonist, NECA, resulted in agonist activity against the three adenosine receptor subtypes, with the following rank of potencies A₁R = A_{2A}R > A_{2B}R (Figure 35A. and Table 8). In contrast, stimulation with the A_{2A}R selective agonist, CGS 21680, solely activated the A_{2A}R significantly (Figure 35B. and Table 8). Stimulating the yeast strains with compounds **36-38** selectively activated the A_{2A}R (Figure 35C. D. E. and Table 8). For compounds **39** and **40**, despite their ability to activate all three adenosine receptors, they were more potent against the A_{2A}R over the other subtypes (Figure 35F. G. and Table 8). Compound **41** failed to elicit a response against any of the adenosine receptors in yeast (Figure 35H. and Table 8). As for A₃R, this receptor cannot be functionally expressed in yeast,[400] thus CHO-K1 cells stably expressing the A₃R (CHO-K1-A₃R) were used. Given that A₃R is G α i/o-coupled, CHO-K1-A₃R cells were co-stimulated with 1 μ M Forskolin (to elevate cAMP) in tandem with each compound to measure its efficacy in inhibiting cAMP. Inhibition of cAMP was only observed upon stimulation with NECA, CGS 21680 and compound **39** (Figure 35 and Table 8). The yeast and mammalian screening confirmed agonist activity of the triazoloquinazolines **36-40** against the adenosine receptor subtypes, exhibiting more potency towards the A_{2A}R, and in particular compounds **36-38** were identified as selective A_{2A}R agonists.

Table 8. Potency (pEC₅₀) and E_{max} values for NECA, CGS 21680 and the triazoloquinazolines **36-41** against A₁, A_{2A} and A_{2B} receptors in GPA1/Gα_{i1/2} or GPA1/Gα_s yeast transplants and the A₃ receptor in CHO-K1-A₃R cells.

	A ₁ R - GPA1/Gα _{i1/2}		A _{2A} R - GPA1/Gα _s		A _{2B} R - GPA1/Gα _s		CHO-K1-A ₃ R	
	pEC ₅₀ ^a	E _{max} ^b	pEC ₅₀ ^a	E _{max} ^b	pEC ₅₀ ^a	E _{max} ^b	pEC ₅₀ ^a	Range ^c
NECA	5.9±0.1	100.4±3.2***	6.3±0.2***	91.1±4.8***	4.2±0.1*	99.8±13.2***	9.5±0.2***	-44.2±2.7***
CGS 21680	ND	ND	4.8±0.2**	107.2±12.2*	ND	ND	7.6±0.2***	-52.9±2.4***
36	NR	NR	5.3±0.2***	69.0±1.9***	NR	NR	NR	NR
37	NR	NR	5.9±0.2***	54.3±5.1***	NR	NR	NR	NR
38	NR	NR	5.8±0.2***	60.2±5.5***	NR	NR	NR	NR
39	5.4±0.3	60.9±9.0***	6.1±0.5***	38.8±5.3***	4.8±0.2*	17.6±2.5***	9.4±0.1***	-17.0±0.9***
40	5.0±0.6	22.7±8.1***	8.3±0.9***	10.0±4.0***	4.6±0.3*	08.7±2.2***	NR	NR
41	NR	NR	NR	NR	NR	NR	NR	NR

Data ± SEM of 4-6 individual replicates

^a Negative logarithm of agonist concentration producing half-maximal response

^b Maximal response observed upon agonist stimulation, as a percentage of that observed upon stimulation with 100 μM NECA

^c Range of response, as a percentage of that observed upon stimulation with 100 μM Forskolin

ND – Not determined – full dose-response curve not feasible

NR – No response

Statistical difference between each agonist and NECA was calculated using a one-way ANOVA with Dunnett's post-test (*, p < 0.05, **, p < 0.01, ***, p < 0.001)

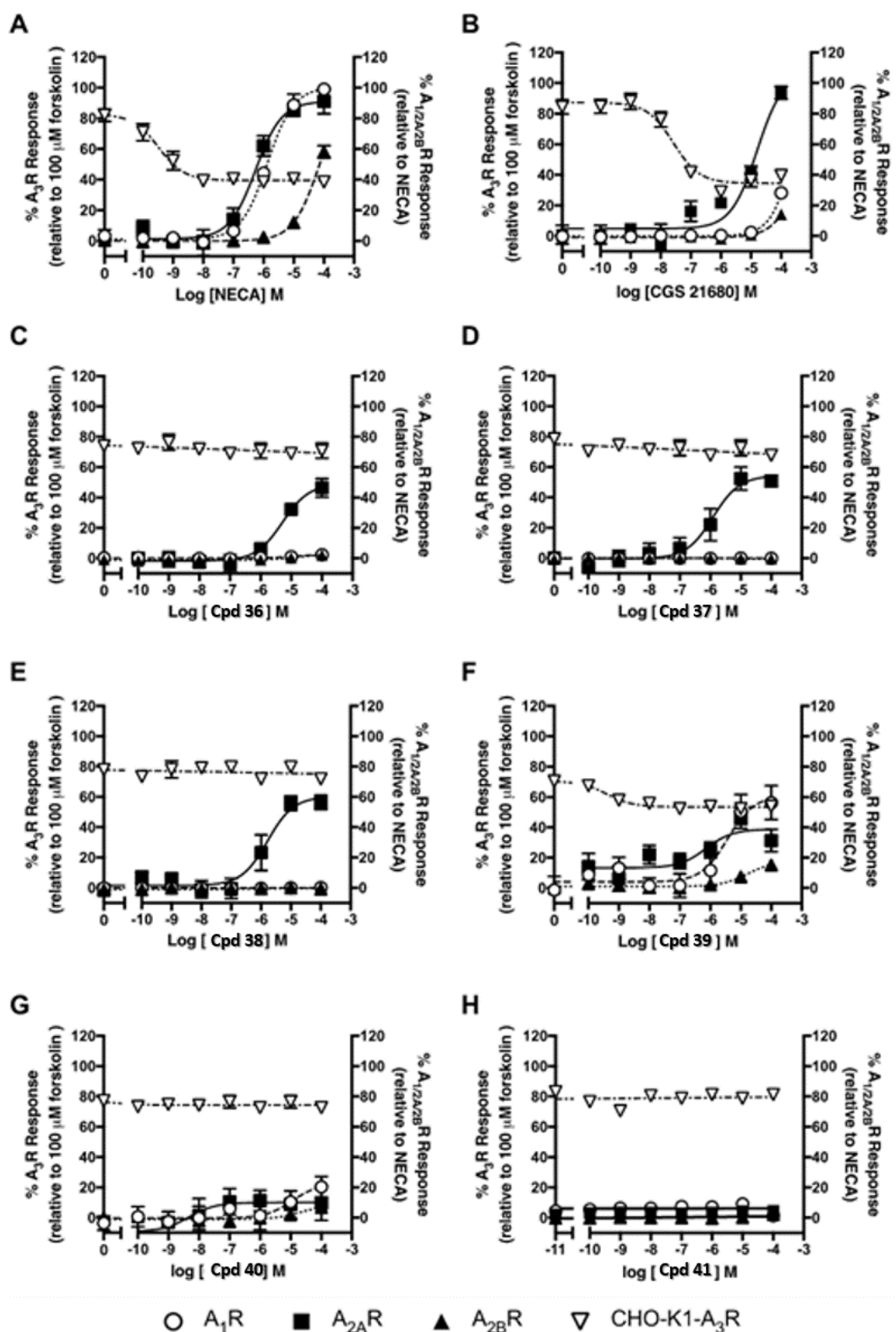


Figure 35. Dose-response curves (produced by the Ladds group members) for NECA, CGS 21680 and compounds **36-41** in either the A_1R and $GPA1/G\alpha_{i1/2}$, $A_{2A}R$ and $GPA1/G\alpha_s$, or the $A_{2B}R$ (with $GPA1/G\alpha_s$ expressed in yeast strains). The efficacy of the compounds (**36-41**) was measured against A_3R in CHO-K1- A_3R cells. Reporter gene activity in yeast was determined using β -galactosidase assays, after 16-hours stimulation with either: NECA (A), CGS 21680 (B) compound **36** (C), compound **37** (D), compound **38** (E), compound **39** (F), compound **40**

(F), compound **41** (G), whereas cAMP inhibition was determined when in CHO-K1-A₃R cells which were co-stimulated with each of the compounds **36-41** and 1 μ M Forskolin. In general, the triazoloquinazolines **36-41** exhibited agonistic activity against the adenosine receptor subtypes, with compounds **36-38** being selective A_{2A}R agonists. The data is represented as either percentage of the response obtained upon stimulating each receptor (A₁R, A_{2A}R, or A_{2B}R) with NECA stimulation, or as a percentage response relative to 100 μ M Forskolin stimulation in the A₃R \pm SEM of 4-6 individual replicates.

4.3.7 Mammalian validation

Despite the fact that the yeast system is often an appropriate tool for screening compounds against the adenosine receptor subtypes, potencies and efficacies may not strongly correlate with those of the mammalian cell types.[434, 437] CHO-K1 cells stably expressing the human A_{2A}R (CHO-K1-A_{2A}R) were used to assess the efficacies of compounds **36-41** in elevating cAMP. In comparison to CGS 21680, the compounds were partial A_{2A}R agonists with a rank order of potency, CGS 21680 > compound **40** = compound **39** > compound **36** = compound **38** > compound **37** > compound **41** (Figure 36 and Table 9). Compounds **36-40**, which are agonists were competitively antagonized by the antagonist ZM 241385 (which in fact is an inverse agonist, this can be seen from the reduction in basal activity) (Figure 36 and Table 9), confirming their agonist activity at the A_{2A}R. As in the case of compound **41**, there was not any competitive antagonism from the ZM 241385 (Figure 36 and Table 9). The shape of the curves obtained in the case of compound **41** could be attributed to the inverse agonism of ZM 241385. In addition, all compounds **36-41** elevated cAMP when 1 μ M ZM 241385 was added or in the absence of A_{2A}R.[319] It was hypothesized that elevated cAMP might be due to the presence of endogenous PDE10A expressed in CHO-K1 cells, and this has been subsequently confirmed by RT-PCR (see Appendix E). By comparing the observed agonist activities of compounds **36-41** upon treatment of CHO-K1 and CHO-K1-A_{2A}R cells, a significant increase in their efficacies was observed in the CHO-K1-A_{2A}R cells (the over-expressed cell line). This may be explained by their ability to stimulate the A_{2A}R, stably expressed in CHO-K1-A_{2A}R cells (Figure 36 and Table 9). For compound **41**, no change in efficacy was obtained suggesting that the elevation of cAMP is solely attributed to the action of the compounds upon other targets in the background of the CHO-K1 cell, potentially PDE10A (which from RT-PCR experiments on the cell system used showed expression of PDE10A). Hence, compounds **36-38** were shown to be selective A_{2A}R agonists, whereas compounds **39-40** are non-selective with limited activity against other adenosine receptors. Some of the observations suggest off-target activity, so a future direction would be to screen the compounds against a panel of receptors commonly expressed in this system. This would also further confirm selectivity for the desired targets.

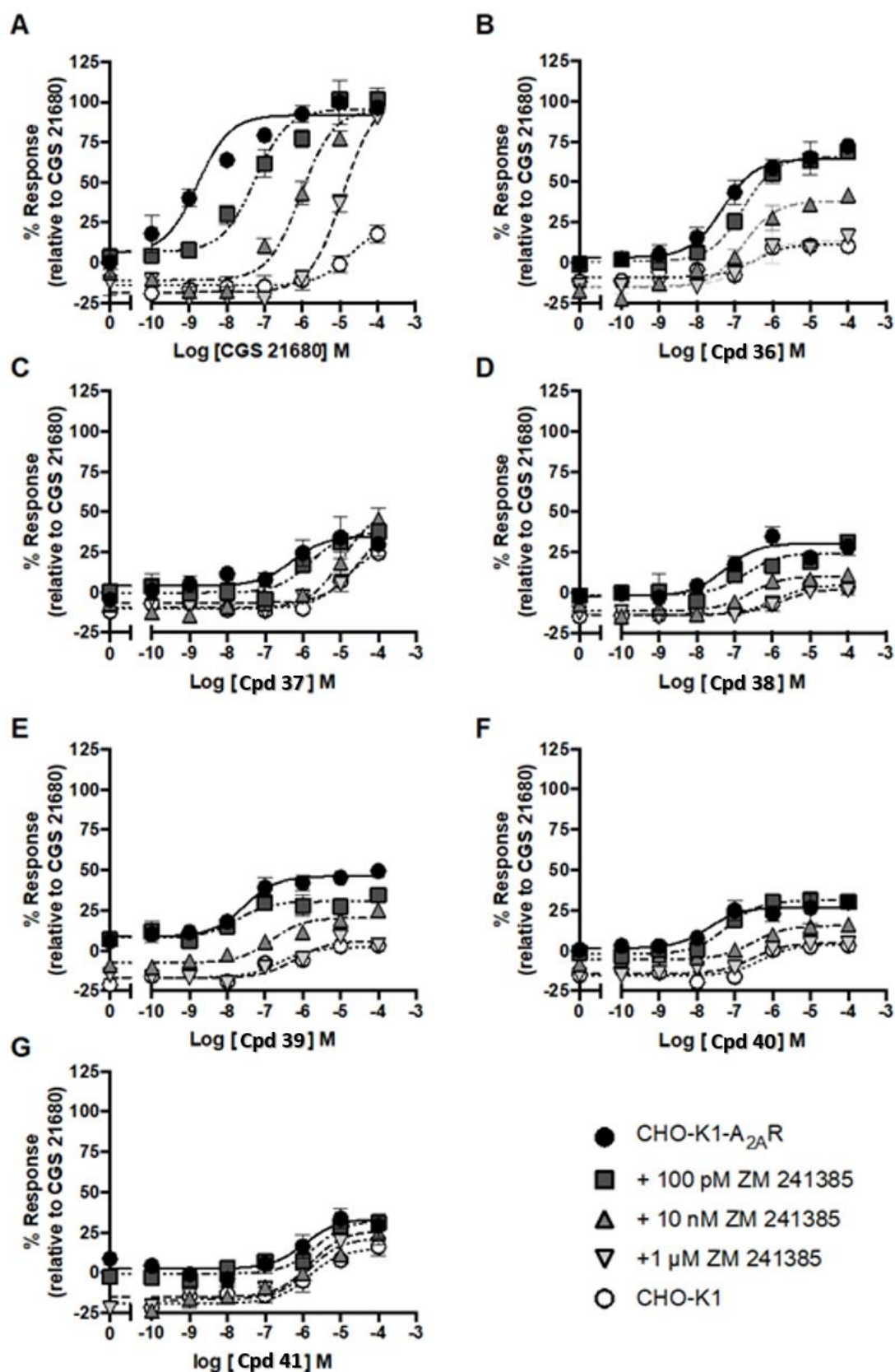


Figure 36. CGS 21680 and compounds **36-41** elevated cAMP in A_{2A}R stably expressed in CHO-K1 cells, which were antagonized by ZM 241385. A_{2A}R stably expressed in CHO-K1 cells (CHO-K1-A_{2A}R) were stimulated for 30 minutes with: CGS 21680 (A), compound **36**

(B), compound **37** (C), compound **38** (D), compound **39** (E), compound **40** (F), or compound **41** (G), after which the cAMP levels were determined. Subsequently compounds were antagonized with either 100 pM, 10 nM or 1 μ M ZM 241385, which decreased the cAMP levels to the same level of CHO-K1 cells (no A_{2A}R stably expressed). Data represented are relative to the response of CGS 21680, \pm SEM of 4-9 individual replicates (plots generated by the Ladds group members).

Table 9. Potency (pEC₅₀) and range of responses for cAMP elevation upon stimulating CHO-K1-A_{2A}R and CHO-K1 cells with CGS 21680 and triazoloquinazolines **36-41**

	CHO-K1-A _{2A} R			CHO-K1			CHO-K1-A _{2A} R vs CHO-K1	
	pEC ₅₀ ^a	Range ^b	N	pEC ₅₀ ^a	Range ^b	n	Δ pEC ₅₀ ^c	Δ Range ^d
CGS 21680	8.8 \pm 0.2**	86.3 \pm 7.2*	9	ND	ND	4	-	-
36	7.3 \pm 0.2***	61.1 \pm 5.2***	8	6.5 \pm 0.3	20.2 \pm 2.7	4	0.8 \pm 0.5	41.0 \pm 7.8
37	6.3 \pm 0.5***	30.5 \pm 8.1***	6	4.9 \pm 0.2	39.5 \pm 3.9	4	1.4 \pm 0.6	-9.0 \pm 10.6
38	7.3 \pm 0.3***	29.0 \pm 6.3***	6	5.9 \pm 0.3	18.3 \pm 2.3	4	1.2 \pm 0.5	10.6 \pm 8.7
39	7.6 \pm 0.2***	37.7 \pm 2.9***	5	6.6 \pm 0.2	18.8 \pm 1.7	4	0.9 \pm 0.6	19.0 \pm 3.6
40	7.7 \pm 0.4***	27.4 \pm 4.4*	6	6.3 \pm 0.2	19.5 \pm 1.7	4	1.3 \pm 2.4	7.9 \pm 6.1
41	6.0 \pm 0.3***	31.1 \pm 5.7**	6	5.8 \pm 0.4	33.9 \pm 5.9	4	0.1 \pm 0.6	-2.9 \pm 8.5

Data \pm SEM of *n* individual replicates

^a Negative logarithm of agonist concentration producing half-maximal response

^b Percentage range of response observed upon agonist stimulation, relative to that obtained with CGS 21680 stimulation in each cell type.

^c Change in pEC₅₀ between CHO-K1 and CHO-K1-A_{2A}R cells (Δ pEC₅₀ = pEC₅₀(CHO-K1-A_{2A}R) - pEC₅₀(CHO-K1))

^d Change in range between CHO-K1 and CHO-K1-A_{2A}R cells (Δ Range = Range (CHO-K1-A_{2A}R) - Range(CHO-K1))

ND – Not determined, full dose-response curve not feasible

Statistical difference, between CHO-K1-A_{2A}R cells and CHO-K1 cells, was calculated using pair-wise t-tests, for each agonist (*, *p* < 0.05, **, *p* < 0.01, ***, *p* < 0.001)

4.3.8 Dual PDE10A inhibition and A_{2A}R agonism is anti-proliferative in CHO-K1-A_{2A}R cells

cAMP elevation has been previously suggested to have *anti-proliferative* effects.[101, 104, 438] Indeed, upon stimulating both of the CHO-K1 and CHO-K1-A_{2A}R cells with Forskolin, inhibition of cell proliferation was observed in a concentration-dependent manner (Figure 37 and Table 10). However, there were not any *anti-proliferative* effects observed upon stimulating the CHO-K1 and CHO-K1-A_{2A}R cells with the selective A_{2A}R agonist, CGS 21680. This

suggests that the activation of the A_{2A}R has no effect upon proliferation (Figure 37 and Table 10). In contrast, compound **36** inhibited CHO-K1-A_{2A}R cells, and compounds **38-40** inhibited the proliferation in CHO-K1 cells, but their pIC₅₀ and I_{max} increased when A_{2A}R was overexpressed in CHO-K1-A_{2A}R cells (Figure 37 and Table 10). Compound **37** displayed anti-proliferative effects in both cell types, which suggests that it might be toxic, whereas compound **41** did not exhibit any anti-proliferative effects upon both cell types, which could imply that the inhibition of PDE10A has no effect upon proliferation (Figure 37 and Table 10).

Table 10. Potency (pIC₅₀) and I_{max} of the responses for the anti-proliferative effects obtained upon simulating by Forskolin, CGS 21680 and triazoloquinazolines **36-41** in CHO-K1-A_{2A}R and CHO-K1 cells

	CHO-K1-A _{2A} R			CHO-K1			CHO-K1-A _{2A} R vs CHO-K1	
	pIC ₅₀ ^a	I _{max} ^b	N	pIC ₅₀ ^a	I _{max} ^b	N	Δ pIC ₅₀ ^c	Δ I _{max} ^d
Forskolin	4.4±0.1***	59.0±2.2***	6	4.5±0.1	62.6±3.4	6	-0.1±0.1	-3.6±4.1
CGS 21680	NR	-2.4±1.4***	4	NR	4.3±10.0	4	NR	-6.7±10.1
36	4.4±0.1***	56.3±5.3***	6	NR	7.9±1.6	6	-	48.5±5.1
37	4.7±0.1****	74.8±2.4***	6	4.8±0.1	78.5±1.4	6	-0.1±0.1	-3.7±2.8
38	4.8±0.1***	70.9±1.8***	6	4.1±0.1	38.0±1.0	6	0.7±0.1	32.9±2.1
39	4.4±0.1****	51.9±0.6***	6	3.5±0.2	19.5±3.7	6	0.9±0.2	32.4±2.1
40	4.7±0.1***	73.2±2.1***	6	3.7±0.1	20.7±1.4	6	1.1±0.1	52.5±2.6
41	NR	-0.4±2.7***	6	NR	4.3±1.9	6	NR	-4.7±3.3

Data ± SEM of *n* individual replicates

^a Negative logarithm of agonist concentration producing half-maximal inhibition

^b Maximal level of inhibition obtained when cells were stimulated with 10 μM agonist relative to that obtained with 1% DMSO treatment

^c Change in pIC₅₀ between CHO-K1 and CHO-K1-A_{2A}R cells (Δ pIC₅₀ = pIC₅₀(CHO-K1-A_{2A}R) - pIC₅₀(CHO-K1))

^d Change in I_{max} between CHO-K1 and CHO-K1-A_{2A}R cells (Δ I_{max} = I_{max} (CHO-K1-A_{2A}R) - I_{max} (CHO-K1))

NR – No response observed

Statistical difference, between CHO-K1-A_{2A}R cells and CHO-K1 cells, was calculated using pair-wise t-tests, for each agonist (*, *p* < 0.05, **, *p* < 0.01, ***, *p* < 0.001)

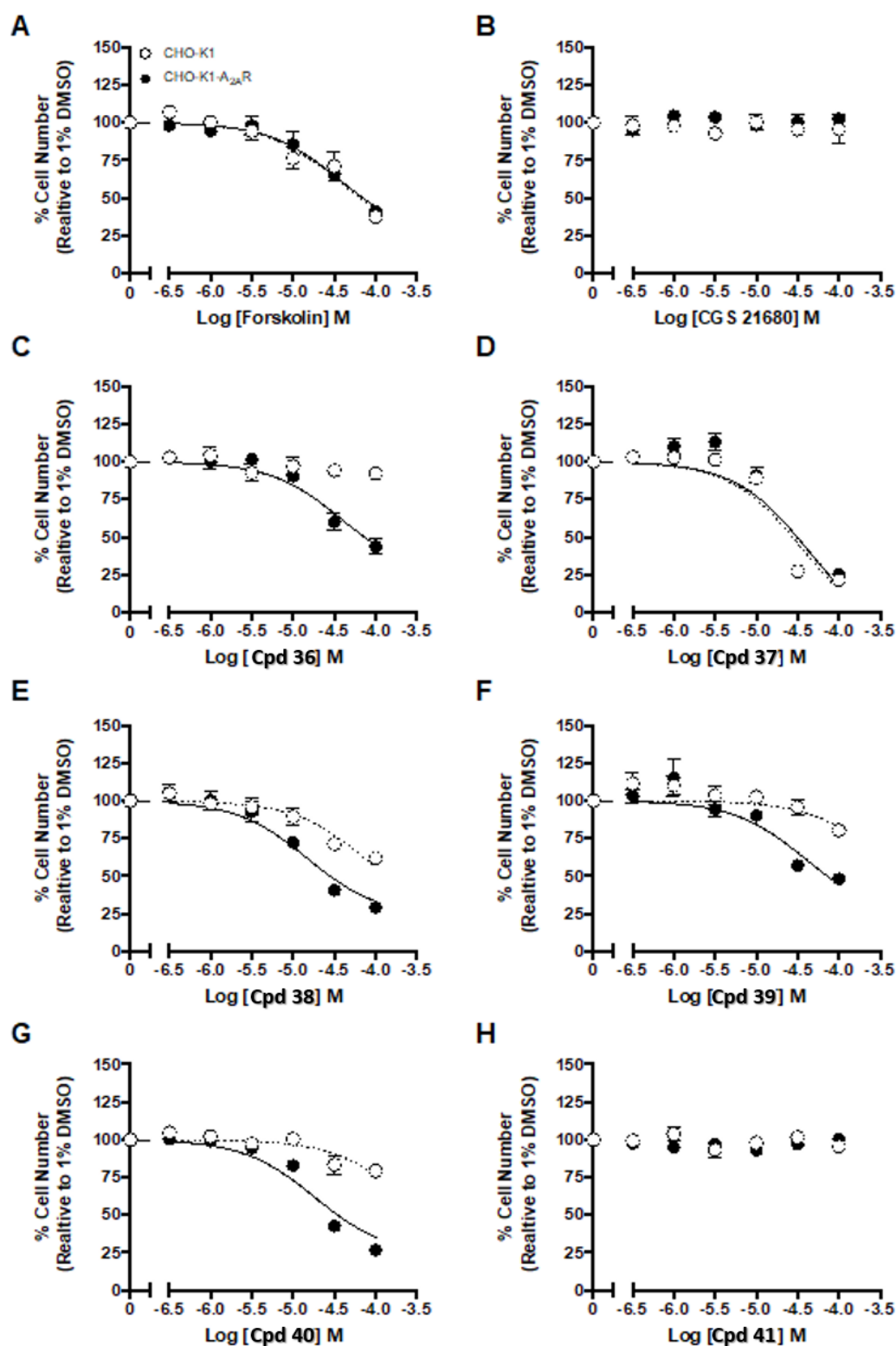


Figure 37. Stimulation by Forskolin and compounds **36**, and **38-40** showed anti-proliferative effects in CHO-K1 cells, which was enhanced when the A_{2A}R is stably expressed in CHO-K1-A_{2A}R cells. As for compound **37**, it displayed anti-proliferative effects in both cell types, which suggests that it might be toxic.

CHO-K1 or CHO-K1-A_{2A}R cells were seeded into 96 well plates and cultured for 24 hours before being stimulated with compounds for 72 hours: Forskolin (*n* = 6) (A), CGS 21680 (*n* = 4) (B), compound **36** (*n* = 6) (C), compound **37** (*n* = 6) (D), compound **38** (*n* = 6) (E), compound **39** (*n* = 6) (F), compound **40** (*n* = 6) (G) or compound **41** (*n* = 6) (H), where CCK-8 was used to determine cell number. Data is represented relative to the cell number obtained upon treatment with 1% DMSO, \pm SEM (plots are produced by the Ladds group members).

4.3.9 Dual PDE10A inhibition and A_{2A}R agonism is anti-proliferative in Lung carcinoma cell-lines

A series of lung carcinoma cell-lines, two lung squamous cell carcinomas (LUSC): LK-2 and H520, and two lung adenocarcinoma cells (LUAC): H1563 and H1792 were employed to ascertain whether the dual PDE10A inhibition and A_{2A}R agonism is anti-proliferative in these cell-lines. First, the target gene expression was quantified for all four adenosine receptors and PDE10A in each cell-line. Then, compounds **36-41** were tested for their ability to elevate cAMP, and their anti-proliferative effects in the four cell-lines were investigated. As for compound **37**, it was not used in these experiments due to its toxicity in CHO-K1 cells but it did exhibit anti-proliferative effects in the four lung carcinoma cell-lines comparable to those observed in CHO-K1 cells.

Low expression of PDE10A and the absence of A_{2A}R were observed in the LK-2 cell-line, while the A₁R, and A_{2B}R are significantly expressed. (Figure 38A. and Table 11). CGS 21680 and compounds **36** and **38-41** were tested for their ability to elevate cAMP, and were found to increase the cAMP levels above basal in a similar manner, (Figure 38A., Table 11). The observed activity of CGS 21680 might be attributed to its action *via* A_{2B}R, whereas the action of compounds **36** and **38-41** might be due to inhibition of PDE10A. Compounds **36** and **38-41**, CGS 21680, and Forskolin were further tested for their anti-proliferative effects in the LK-2 cell-line. Inhibition of cell growth was only observed for Forskolin and compound **38** (Figure 38A. and Table 12). This could be attributed to the fact that compound **38** was the most potent in elevating cAMP in the LK-2 cell-line.

In the H520 cell-line, all of four adenosine receptor sub-types were found to be expressed, and the PDE10A was also present at low concentrations (Figure 38B.). With the exception of compounds **40-41**, all tested compounds were more potent in elevating cAMP in comparison to their response in the LK-2 cell-line (Figure 38B. and Table 11). Increased anti-proliferative effects were observed in tandem with cAMP elevation obtained in this cell-line, in particular for compounds **36** and **38-39** (Figure 38B. and Table 12). In the H1792 cell-line an increase in

PDE10A expression was detected in comparison to the H520 cell-line, and all of the four adenosine receptors were expressed (Figure 38C.). Similar to what was observed in the LK-2 and H520 cell-lines, compounds **36** and **38-40** elevated cAMP levels, and they exhibited anti-proliferative effects. In particular, compound **39** was the most potent among the tested compounds in both cAMP elevation (Table 11) and anti-proliferation (Table 12) in the H1792 cell-line, compared to its response in the LK-2 and H520 cell-line. This could be attributed to the increased combined expression of PDE10A and adenosine receptor subtypes in this cell-line (Figure 38C.).

In the H1563 cell-line, there was the highest combined target expression of A_{2A}R and PDE10A, which could explain the highest sensitivity to anti-proliferation observed among the four cell-lines (Figure 38D. and Table 12). In particular, compound **36**, a selective A_{2A}R agonist and a PDE10A inhibitor, exhibited a strong correlation between its pEC₅₀ (4.91±0.2, 5.57±0.1, 5.98±0.1, 6.42±0.1 respectively) and pIC₅₀ (NR, 3.34±0.10, 4.00±0.07, 4.37±0.04 respectively) values. They both increased along with the levels of combined target expression (A_{2A}R and PDE10A) across cell lines, from LK-2 – H520 – H1792 – H1563.

As a general trend, an increase was observed for pIC₅₀ values for compounds **36**, **38-40**, as the combined target expression (A_{2A}R and PDE10A) increased across cell lines, from LK-2 – H520 – H1792 – H1563. The least sensitive was LK-2, expressing a small level of PDE10A, and the most sensitive was H1563, expressing similar levels of A_{2A}R to that of H1792. However, PDE10A is over expressed in comparison to the other three cell types (Figure 38). The increase in potency (pIC₅₀) (Figure 39A. and Table 12) and I_{max} was most remarkable for compounds **36** and **40** (Figure 39B. and Table 12). As for the increase in cAMP elevation (pEC₅₀ values), this also correlated with the increase in the combined expression of A_{2A}R and PDE10A, being most prominent for compounds **36** and **38** (Figure 39C. and Table 11). For compound **41**, it did not exhibit any anti-proliferative effects in any of the four cell-lines, but rather an increase in proliferation was observed. This could be attributed to the fact that the anti-proliferative effects of PDE10A inhibition might be specific to the cancer cell type. For CGS 21680, it was only anti-proliferative in the H1563 cell-line, which might be attributed to the high sensitivity of the cell-line to anti-proliferation. Hence, it appears that the A_{2A}R activation alone weakly exhibits anti-proliferative effects, whereas it is not clear from the experiment whether sole inhibition of PDE10A could increase proliferation or have no effects. In contrast, PDE10A inhibition has, for instance, been reported to be anti-proliferative on other cancer cells,[439] and agonsim of A_{2A}R was shown to increase fibroblast proliferation.[440] Therefore, it could be inferred that

neither A_{2A}R agonists nor PDE inhibitors may solely induce therapeutically relevant cAMP mediated cell death in these particular cell-lines.

It was noted that the maximum achievable IC₅₀ values for the compounds showing anti-proliferative effects on lung carcinoma cell lines were those with activities of 30-50 μ M, despite the increase in PDE10A expression in the most sensitive cell types (Figure 39 and Table 12). This may be due to protocol, which includes chronic stimulation by exposing cells to each compound for 72 hours. Indeed, it has been previously reported that GPCRs can desensitize and/or internalize upon short or long-term stimulation with agonists[441, 442] which may be independent of their potency. This is generally considered to be due to β -arrestin binding followed by disposal of the receptor complex in clathrate-coated pits. There is also signaling associated with this process. Biased signaling or functional selectivity is not considered in this work, although this would be an interesting topic for further exploration[442, 443]. This suggests that targeting G α s-coupled receptors alone as an anti-proliferative strategy may be of limited benefit.

Additionally, when comparing the pIC₅₀ and pEC₅₀ values obtained for proliferation and cAMP elevation across the CHO-K1-A_{2A}R, LK-2, H520, H1792, H1563 cell-lines, a strong correlation (Figure 39D. $r = 0.8238$, 95% confidence interval (CI); 0.6232 – 0.9227) is obtained for all the anti-proliferative compounds. This finding appears to be more consistent than in previous studies, where it was found that the effect of cAMP elevation is cell type specific, [99–106] and the magnitude of cAMP elevation did not show a strong correlation with the levels of anti-proliferative action.[438] Hence, this study suggests a causative link between cAMP elevation and the anti-proliferative effects observed, which is further supported by the anti-proliferative effects exhibited by Forskolin (Figure 39 and Table 12). This is similar to what was observed in other cell types where increased cAMP levels (induced by Forskolin) inhibited fibroblast proliferation.[440, 444]

Particularly in the case of compound **36**, it has been shown that the increased ability to elevate cAMP levels through dual A_{2A}R activation and inhibition of PDE10A results in greater anti-proliferative effects on lung carcinoma cell-lines. This finding suggests that by improving efficacy of compounds for cAMP elevation, this would in turn increase the efficacy of their anti-proliferative effects in lung carcinoma. Indeed, it has been reported that sustained cAMP elevation can impact gene expression.[445] In addition, this is in agreement with what has been reported in the literature, namely that A_{2A}R agonists in combination with various PDE

inhibitors were anti-proliferative.[98] The novelty of this approach, however, lies in combining this strategy into a single ligand, which results in the same anti-proliferative effects with potencies which are comparable to those of cisplatin on H520, H1563 and H1792 cell-lines.[446] Thus, there is the possibility that targeting both of these proteins may have therapeutic potential in various carcinomas, particularly lung adenocarcinoma.[130–132]

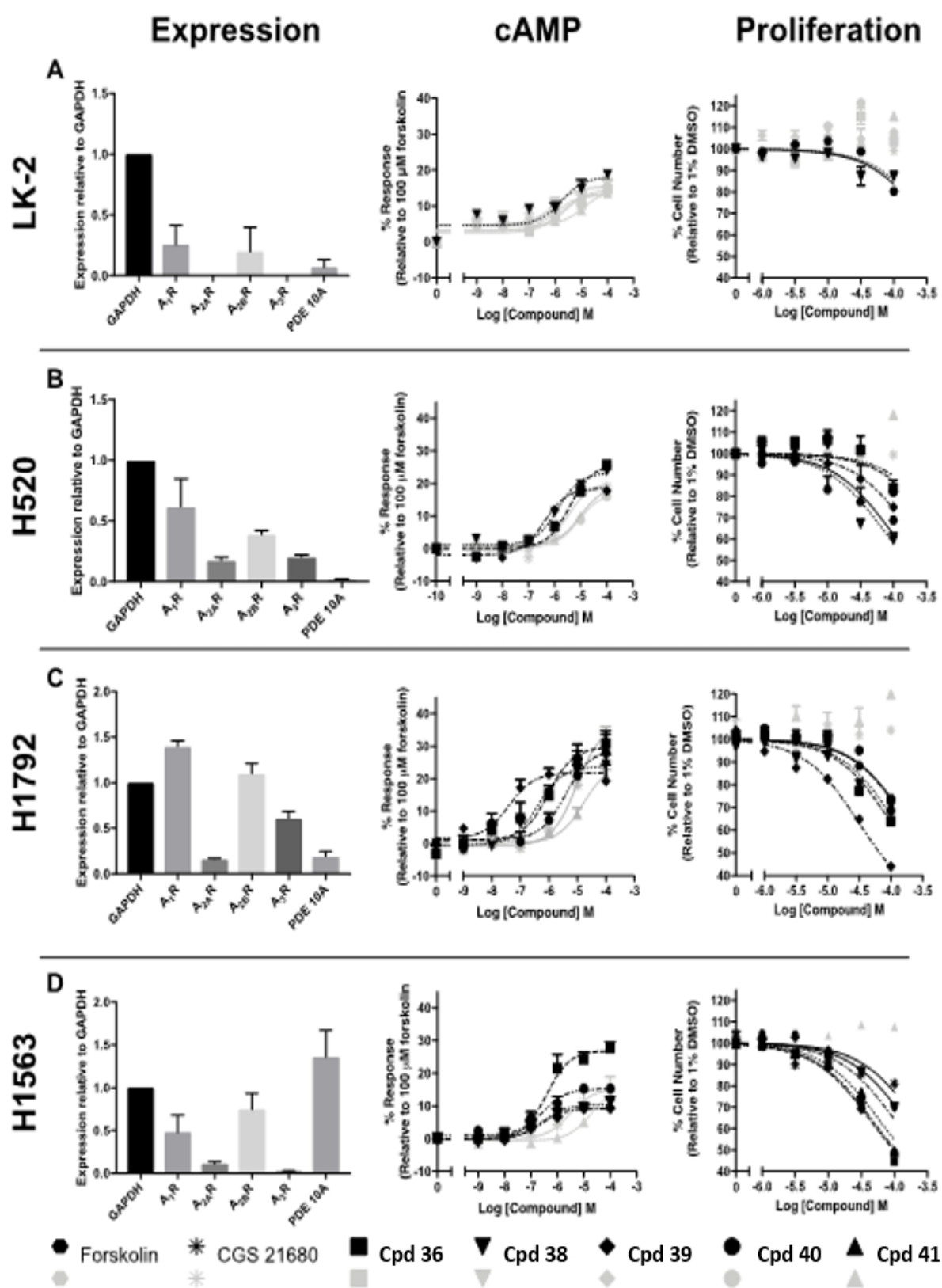


Figure 38. Lung squamous cell carcinoma and adenocarcinoma cells show increased sensitivity in terms of proliferation to triazoloquinazolines **36-41**, depending on the combined expression of A₂A_R and PDE10A. RT-PCR analysis was performed (by DS) on Lung squamous cell carcinoma cells; LK-2 (A), and H520 (B), and lung adenocarcinoma cells: H1792 (C) and

H1563 (D) in order to determine expression of the A₁R, A_{2A}R, A_{2B}R, A₃R and PDE10A. Data are represented in relation to GAPDH expression \pm SEM of 3 individual replicates. Subsequently, stimulation of each cell line with CGS 21680 or compounds **36**, **38-41** was performed for 30 minutes, after which cAMP levels were measured. Data are relative to the response of 100 μ M Forskolin, \pm SEM of 4-8 individual replicates. In addition, stimulation of all cells with CGS 21680, or compounds **36**, **38-41** for 72 hours was performed, after which their cell number was determined with CCK-8. Data is relative to the cell number obtained after treatment with 1% DMSO, \pm SEM of 4-8 replicates.

Table 11. Potency (pEC₅₀) and E_{max} values for cAMP accumulation of CGS 21680 and triazoloquinazolines **36**, **38-41** stimulated by LK-2, H520, H1792 and H1563 lung carcinoma cell-lines.

	LK-2		H520		H1792		H1563	
	pEC ₅₀ ^a	E _{max} ^b	pEC ₅₀ ^a	E _{max} ^b	pEC ₅₀ ^a	E _{max} ^b	pEC ₅₀ ^a	E _{max} ^b
CGS 21680	5.1 \pm 0.1	17.1 \pm 0.9	5.5 \pm 0.2**	21.3 \pm 1.0	5.5 \pm 0.1***	27.3 \pm 2.1	5.3 \pm 0.1***	15.5 \pm 1.6***
36	4.9 \pm 0.2	18.4 \pm 1.9	5.6 \pm 0.1*	26.1 \pm 3.0	6.0 \pm 0.1***	32.2 \pm 2.8	6.4 \pm 0.1***	29.0 \pm 1.1***
38	5.5 \pm 0.2	18.30 \pm 1.5	5.9 \pm 0.2**	23.4 \pm 1.4	6.5 \pm 0.1***	23.7 \pm 1.5	6.5 \pm 0.2***	12.5 \pm 0.7***
39	5.3 \pm 0.2	14.3 \pm 2.3	6.4 \pm 0.3**	18.5 \pm 1.8	7.5 \pm 0.1***	21.8 \pm 1.8	6.5 \pm 0.1***	10.0 \pm 0.6***
40	5.5 \pm 0.2	15.2 \pm 1.3	5.3 \pm 0.1**	16.6 \pm 2.7	5.6 \pm 0.2***	27.8 \pm 3.1	6.8 \pm 0.2***	15.7 \pm 0.8***
41	5.0 \pm 0.2	14.7 \pm 1.7	5.2 \pm 0.1**	23.1 \pm 2.7	4.3 \pm 0.1***	24.9 \pm 2.5	4.7 \pm 0.2***	15.8 \pm 1.9***

Data \pm SEM of 4-8 individual replicates

^a Negative logarithm of agonist concentration producing half-maximal response

^b Maximal response observed upon agonist stimulation, as a percentage of that observed upon stimulation with 100 μ M Forskolin

The statistical difference between each agonist and CGS 21680 was calculated using a one-way ANOVA with Dunnett's post-test (*, p < 0.05, **, p < 0.01, ***, p < 0.001)

Table 12. Anti-proliferative effects of CGS 21680, Forskolin and triazoloquinazolines **36**, **38-41** stimulated BY LK-2, H520, H1792 and H1563 lung carcinoma cell-lines

	LK-2		H520		H1792		H1563	
	pIC ₅₀ ^a	I _{max} ^b	pIC ₅₀ ^a	I _{max} ^b	pIC ₅₀ ^a	I _{max} ^b	pIC ₅₀ ^a	I _{max} ^b
CGS 21680	NR	-3.6 \pm 2.6	NR	0.6 \pm 2.0***	NR	-4.1 \pm 1.6***	3.7 \pm 0.1	18.9 \pm 1.8***
Forskolin	3.5 \pm 0.1	-19.7 \pm 2.0	4.1 \pm 0.1	31.2 \pm 2.1***	3.7 \pm 0.1	25.9 \pm 1.0***	3.8 \pm 0.1	29.9 \pm 1.1***
36	NR	-3.6 \pm 0.7	3.3 \pm 0.1	16.0 \pm 3.8***	4.0 \pm 0.1	36.0 \pm 2.7***	4.4 \pm 0.1	55.3 \pm 1.5***
38	3.4 \pm 0.1	12.6 \pm 1.8	4.2 \pm 0.1	40.2 \pm 2.9***	3.9 \pm 0.1	29.4 \pm 2.8***	4.2 \pm 0.1	49.3 \pm 3.7***
39	NR	-5.2 \pm 1.9	3.8 \pm 0.1	25.0 \pm 1.5***	4.5 \pm 0.1	56.0 \pm 0.9***	4.0 \pm 0.1	29.9 \pm 2.9***
40	NR	-7.3 \pm 1.9	3.2 \pm 0.3	17.9 \pm 1.0***	3.8 \pm 0.1	31.4 \pm 2.0***	4.3 \pm 0.1	51.0 \pm 0.5***
41	NR	-15.2 \pm 2.1	NR	-18.2 \pm 3.1***	NR	-19.8 \pm 5.8***	NR	-7.8 \pm 2.1***

Data \pm SEM of 4-8 individual replicates

^a Negative logarithm of agonist concentration producing half-maximal inhibition

^b Maximal level of inhibition obtained when cells were stimulated with 10 μ M agonist relative to that obtained with 1% DMSO treatment

Statistical difference for each agonist compared to its effect upon LK-2 cells was calculated using a one-way ANOVA with Dunnett's post-test (*, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$)

Figure 39. Bar and scatter plots show the relative effects of CGS 21680, Forskolin, or

Figure 39. (continued) compounds **36**, **38-41** in cAMP elevation and proliferation assays for LK-2, H520, H1563 and H1792 cell-lines (produced by the Ladds group members).

A. A summary of the potencies (pIC_{50} values) for compounds **36**, **38-41**, CGS 21680 and Forskolin for inhibition of proliferation of LK-2, H520, H1792 and H1563 cells. Bars denote the mean (\pm SEM) of data points, whereas individual data are represented by scatter plots.

B. An overview of the maximal level of inhibition of proliferation (I_{max}) obtained upon treatment of each of the LK-2, H520, H1792 and H1563 cell-lines with compounds **36**, **38-41**, CGS 21680 and Forskolin. Bars denote the mean I_{max} values \pm SEM, whereas a scatter plot represents individual data points. A one-way ANOVA with Dunnett's post-test was used to assess the statistical significance of the response obtained in LK-2 cell-line.

C. Potencies (pEC_{50} values) obtained for compounds **36**, **38-41**, CGS 21680 and Forskolin in the cAMP accumulation assays upon in LK-2, H520, H1792 and H1563 cell-lines (30 minutes exposure). The mean pEC_{50} values \pm SEM are represented by bars, whereas individual data points are represented by a scatter plot. A one-way ANOVA with Dunnett's post-test was used to assess the statistical significance of the response obtained in LK-2 cell-line.

D. Comparison between the pIC_{50} and pEC_{50} values obtained for proliferation and cAMP elevation of all the anti-proliferative compounds across the CHO-K1- $\text{A}_{2\text{A}}\text{R}$, LK-2, H520, H1792, H1563 cell-lines, shows a strong correlation with $r = 0.8238$, 95% confidence interval (CI); 0.6232 – 0.9227).

4.4 Conclusions

In this work, a novel structure-based approach has been successful in identifying triazoloquinazolines as the first dual ligands that activate the A_{2A}R and inhibit PDE10A simultaneously. Docking of triazoloquinazolines **36-41**, which are known PDE10A inhibitors, was performed on the orthosteric site of A_{2A}R (PDB ID: 2YDO), and it is demonstrated experimentally that these ligands are indeed A_{2A}R agonists. The agonist activity of the compounds was measured by a yeast-screening assay and also in mammalian cells where both assays confirmed that compounds **36-40** were A_{2A}R agonists, and revealed that compounds **36-38** are selective for the A_{2A}R. In addition, the five compounds produced dose-response curves obtained from mammalian cells expressing the A_{2A}R with pEC₅₀ values ranging from 6.3 to 7.7. It is suggested that the observed A_{2A}R sub-type selectivity for **36-38** is attributed to their predicted interactions with the His₂₅₀ residue, which is an interaction present only in the selective co-crystallized A_{2A}R agonists, such as CGS 21680 and UK432097. It was further demonstrated by MD simulation analysis that this residue undergoes conformational changes only when selective A_{2A}R agonists are bound and not when non-selective agonists bind to A_{2A}R. This could contribute to shaping the orthosteric site to favor selectivity of A_{2A}R agonists.

Moreover, MD simulations analysis highlighted the motion of Val₈₄ in TM3 as an essential requirement for A_{2A}R activation. Compounds **36** and **38-40** exhibited concentration-dependent anti-proliferative effects in lung squamous cell carcinoma cells and lung adenocarcinoma cells which correlated with co-expression of A_{2A}R and PDE10A and increased cellular levels of cAMP. In particular, compound **36** (as a selective A_{2A}R agonist and a PDE10A inhibitor) exhibited correlation between its pEC₅₀ and pIC₅₀ values, which increased in tandem with the combined target expression (A_{2A}R and PDE10A) across cell lines, from LK-2 – H520 – H1792 – H1563. Hence, the MD-assisted approach proposed in this work provides a template for generating A_{2A}R agonists as part of a dual-target design objective, which demonstrates promising therapeutic benefits, particularly in lung adenocarcinoma cell-lines.

5 Conclusions & Future Work

Novel single- and multi-target ligand design approaches have been presented in this work, which could be of general use to a wide range of ligand design problems, across (multi-factorial) disease areas and target families. The design approaches investigated here resulted in the discovery of novel compounds with activities at multiple proteins/multiple binding sites including allosteric ligands. First, chemically novel allosteric inhibitors of calpain-1 were identified using the PEF(S) domain, which may offer improved selectivity towards the enzyme. Second, multi-target ligands with high binding affinity at PDE10A, A₁ and A_{2A} receptors (2-aminopyridine-3-carbonitriles) were found, which may in concert elevate intracellular cAMP deemed to be beneficial in neurodegenerative diseases. Third, dual ligands were discovered at the A_{2A}R and PDE10A (triazoloquinazolines), which were successful in elevating cAMP at these targets, and have been further shown by experiments to exhibit potential clinical value in lung carcinoma.

Future studies could be directed towards finding new dual ligands at the A_{2A}R and PDE10A following the structure-based approach described in chapter 4. This could involve docking and MD simulations of compounds that are experimentally known PDE10A inhibitors including other triazoloquinazolines and additionally compounds with new chemical scaffolds. In particular, it would be interesting to investigate the other chemical series that were predicted (in chapter 3)[338] as multi-target ligands against the A_{2A}R and PDE10A by the structure- and ligand- based techniques. A particular focus in the MD simulation analysis could be on the dynamics of Val₈₄ in the A_{2A} receptor upon ligand binding (see chapter 4) in order to shortlist candidates of A_{2A}R agonists. This could be followed by experiments to validate these as agonists,[400] with further investigation of their selectivity profile against other subtypes of both protein families (e.g. A₁R, A_{2B}R, A₃R, PDE7A, PDE7B, and PDE9A).

Furthermore, it would be interesting to explore biased cAMP signaling/functional selectivity[443] of the novel A_{2A}R agonists. Subsequently, the anti-proliferative effects would be tested in lung carcinoma cell-lines. The overall aim of this future study is to search for new dual-target ligands with confirmed selectivity against these targets along with signaling bias towards cAMP, which could enhance their potency (while reducing desensitization)[441] in inhibiting the proliferation of lung carcinoma cell-lines. This could eventually lead to testing against lung cancer models *in vivo*.

Additionally, molecular dynamic analysis could be performed on the allosteric calpain-1 inhibitors that were discovered (chapter 2)[447] once (if) the crystal structure of the full calpain-1 complex becomes available. Co-crystallization of these novel ligands would confirm the binding and ligand-protein interactions proposed in this work. The dynamics would aim to develop an understanding of the causative link between the specific binding of the compounds to the PEF(S) domain and their functional effect i.e. their ability to inhibit the activity of the calpain-1 enzyme. These suggested mechanistic studies could aid in discovering new and more potent allosteric calpain-1 inhibitors, which may offer improved selectivity and a reduced side-effects profile.

Hence, finding selective and potent allosteric calpain-1 inhibitors would open the door towards discovering potential drug candidates. In addition, identifying new compounds that are more potent in inhibiting the proliferation of lung cancer cells could potentially constitute a huge step forward in the treatment of Lung cancer.

References

1. Gleeson MP, Hersey A, Montanari D, Overington J (2011) Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov* 10:197–208
2. Huggins DJ, Sherman W, Tidor B (2012) Rational approaches to improving selectivity in drug design. *J Med Chem* 55:1424–1444
3. Farid R, Day T, Friesner RA, Pearlstein RA (2006) New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorganic Med Chem* 14:3160–3173
4. De Graaf C, Vermeulen NPE, Feenstra KA (2005) Cytochrome P450 in silico: An integrative modeling approach. *J Med Chem* 48:2725–2755
5. Obach RS, Walsky RL, Venkatakrishnan K, et al (2005) The Utility of in Vitro Cytochrome P450 Inhibition Data in the Prediction of Drug-Drug Interactions. *J Pharmacol Exp Ther* 316:336–348
6. Noble MEM, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: insights into drug design from structure. *Science* 303:1800–5
7. Knight ZA, Shokat KM (2005) Features of selective kinase inhibitors. *Chem Biol* 12:621–637
8. Hopkins AL, Mason JS, Overington JP (2006) Can we rationally design promiscuous drugs? *Curr Opin Struct Biol* 16:127–136
9. Lazaridis T (2002) Binding Affinity and Specificity from Computational Studies. *Curr Org Chem* 6:1319–1332
10. Cheng AC (2008) Chapter 2 Predicting Selectivity and Druggability in Drug Discovery. In: *Annual Reports in Computational Chemistry*. pp 23–37
11. Ortiz AR, Gomez-Puertas P, Leo-Macias A, et al (2006) Computational approaches to model ligand selectivity in drug design. *Curr Top Med Chem* 6:41–55
12. Cortés-Ciriano I, Ain QU, Subramanian V, et al (2015) Polypharmacology modelling using proteochemometrics (PCM): Recent methodological developments, applications to target families, and future prospects. *Medchemcomm* 6:24–50
13. Johnstone S, Albert JS (2017) Pharmacological property optimization for allosteric ligands: A medicinal chemistry perspective. *Bioorganic Med Chem Lett* 27:2239–2258
14. Yuan Y, Tam MF, Simplaceanu V, Ho C (2015) A new look at Hemoglobin assembly. *Chem Rev* 115:1702–1724
15. Gan-or Z, Rouleau GA (2016) Calpain 1 in neurodegeneration: a therapeutic target? *Lancet Neurol* 15:1118
16. Ennes-Vidal V, Menna-Barreto RFS, Branquinha MH, et al (2017) Why calpain inhibitors are interesting leading compounds to search for new therapeutic options to treat leishmaniasis? *Parasitology* 144:117–123
17. Auvin S, Pignol B, Navet E, et al (2006) Novel dual inhibitors of calpain and lipid

- peroxidation with enhanced cellular activity. *Bioorganic Med Chem Lett* 16:1586–1589
18. Adams SE, Robinson EJ, Miller DJ, et al (2015) Conformationally restricted calpain inhibitors. *Chem Sci* 6:6865–6871
 19. Wang J, Wang ZM, Li XM, et al (2016) Synthesis and evaluation of multi-target-directed ligands for the treatment of Alzheimer's disease based on the fusion of donepezil and melatonin. *Bioorganic Med Chem* 24:4324–4338
 20. Langhammer S (2013) Rationale for the design of an oncology trial using a generic targeted therapy multi-drug regimen for NSCLC patients without treatment options (Review). *Oncol Rep* 30:1535–1541
 21. Unzeta M, Esteban G, Bolea I, et al (2016) Multi-target directed donepezil-like ligands for Alzheimer's disease. *Front Neurosci* 10:1–24
 22. Więckowska A, Kołaczkowski M, Bucki A, et al (2016) Novel multi-target-directed ligands for Alzheimer's disease: Combining cholinesterase inhibitors and 5-HT₆receptor antagonists. Design, synthesis and biological evaluation. *Eur J Med Chem* 124:63–81
 23. Li K, Schurig-Briccio LA, Feng X, et al (2014) SAR analysis of new anti-TB drugs currently in pre-clinical and clinical development. *J Med Chem* 57:3126–3139
 24. Sikazwe D (2012) The Multi-Target Drug Design Era is Here, Consider it. *Drug Des* 1:1–2
 25. Gazova Z, Soukup O, Sepsova V, et al (2017) Multi-target-directed therapeutic potential of 7-methoxytacrine-adamantylamine heterodimers in the Alzheimer's disease treatment. *Biochim Biophys Acta - Mol Basis Dis* 1863:607–619
 26. Van Der Schyf CJ (2011) The use of multi-target drugs in the treatment of neurodegenerative diseases. *Expert Rev Clin Pharmacol* 4:293–298
 27. Nath O, Singh A, Singh IK (2017) In-Silico Drug discovery approach targeting receptor tyrosine kinase-like orphan receptor 1 for cancer treatment. *Sci Rep* 7:1–10
 28. Maione P (2006) Combining Targeted Therapies and Drugs with Multiple Targets in the Treatment of NSCLC. *Oncologist* 11:274–284
 29. Xie L, Bourne PE (2015) Developing multi-target therapeutics to fine-tune the evolutionary dynamics of the cancer ecosystem. *Front Pharmacol* 6:1–5
 30. Hughes RE, Nikolic K, Ramsay RR (2016) One for all? Hitting multiple Alzheimer's disease targets with one drug. *Front Neurosci* 10:1–10
 31. Sameem B, Saeedi M, Mahdavi M, Shafiee A (2017) A review on tacrine-based scaffolds as multi-target drugs (MTDLs) for Alzheimer's disease. *Eur J Med Chem* 128:332–345
 32. Agis-Torres A, Sollhuber M, Fernandez M, Sanchez-Montero JM (2014) Multi-Target-Directed Ligands and other Therapeutic Strategies in the Search of a Real Solution for Alzheimer's Disease. *Curr Neuropharmacol* 12:2–36
 33. Bansal Y, Silakari O (2014) Multifunctional compounds: Smart molecules for multifactorial diseases. *Eur J Med Chem* 76:31–42

34. Laura M, Cavalli A (2016) Multitarget Drug Discovery and Polypharmacology. *ChemMedChem* 11:1190–1192
35. Ramsay RR, Popovic-Nikolic MR, Nikolic K, et al (2018) A perspective on multi-target drug discovery and design for complex diseases. *Clin Transl Med* 7:1–14
36. Lin HH, Zhang L Le, Yan R, et al (2017) Network Analysis of Drug-target Interactions: A Study on FDA-approved New Molecular Entities between 2000 to 2015. *Sci Rep* 7:12230
37. Lu JJ, Pan W, Hu YJ, Wang YT (2012) Multi-target drugs: The trend of drug research and development. *PLoS One* 7:e40262
38. Cavalli A, Bolognesi ML, Minarini A, et al (2008) Multi-target-directed ligands to combat neurodegenerative diseases. *J Med Chem* 51:347–372
39. Viayna E, Sola I, Di Pietro O, Munoz-Torrero D (2013) Human Disease and Drug Pharmacology, Complex as Real Life. *Curr Med Chem* 20:1623–1634
40. Marzo A, Dal Bo L, Monti NC, et al (2004) Pharmacokinetics and pharmacodynamics of safinamide, a neuroprotectant with antiparkinsonian and anticonvulsant activity. *Pharmacol Res* 50:77–85
41. Maeda K, Sugino H, Akazawa H, et al (2014) Brexpiprazole I: In Vitro and In Vivo Characterization of a Novel Serotonin-Dopamine Activity Modulator. *J Pharmacol Exp Ther* 350:589–604
42. Frankel JS, Schwartz TL (2017) Brexpiprazole and cariprazine: distinguishing two new atypical antipsychotics from the original dopamine stabilizer aripiprazole. *Ther Adv Psychopharmacol* 7:29–41
43. Peters J (2013) Polypharmacology – Foe or Friend ? *J Med Chem* 56:8955–8971
44. Li YH, Wang PP, Li XX, et al (2016) The human kinome targeted by FDA approved multi-target drugs and combination products: A comparative study from the drug-target interaction network perspective. *PLoS One* 11:e0165737
45. Santos R, Ursu O, Gaulton A, et al (2016) A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16:19–34
46. Okamoto K, Kodama K, Takase K, et al (2013) Antitumor activities of the targeted multi-tyrosine kinase inhibitor lenvatinib (E7080) against RET gene fusion-driven tumor models. *Cancer Lett* 340:97–103
47. Tsou H-R, Overbeek-Klumpers EG, Hallett WA, et al (2005) Optimization of 6,7-Disubstituted-4-(arylamino)quinoline-3-carbonitriles as Orally Active, Irreversible Inhibitors of Human Epidermal Growth Factor Receptor-2 Kinase Activity. *J Med Chem* 48:1107–1131
48. Asghar U, Witkiewicz AK, Turner NC, Knudsen ES (2015) The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat Rev Drug Discov* 14:130–146
49. Levis M (2017) Midostaurin approved for FLT3-mutated AML. *Blood* 129:3403–3406
50. Morphy J, Harris C (2012) Designing multi-target drugs. The Royal Society of Chemistry, Cambridge

51. Peters J (2012) Polypharmacology in drug discovery. Wiley, Hoboken
52. Decker M (2017) Design of hybrid molecules for drug development. Elsevier, Amsterdam
53. Multiple target ligand database. <http://www.mtdcadd.com>.
54. DrugBank. <https://www.drugbank.ca>.
55. Therapeutic target database. <http://bidd.nus.edu.sg/group/ttd/ttd.asp>
56. Hopkins AL (2008) Network pharmacology: The next paradigm in drug discovery. *Nat Chem Biol* 4:682–690
57. Bottegoni G, Favia AD, Recanatini M, Cavalli A (2012) The role of fragment-based and computational methods in polypharmacology. *Drug Discov Today* 17:23–34
58. Talevi A (2015) Multi-target pharmacology: possibilities and limitations of the “skeleton key approach” from a. *Front Pharmacol* 6:1–7
59. Morphy R, Rankovic Z (2010) Lead generation approaches in drug discovery. In: Design of multitarget ligands. Wiley, Hoboken, pp 141–164
60. Morphy R, Rankovic Z (2009) Designing Multiple Ligands – Medicinal Chemistry Strategies and Challenges. *Curr Pharm Des* 15:587–600
61. Morphy R, Rankovic Z (2005) Designed Multiple Ligands. An Emerging Drug Discovery Paradigm. *J Med Chem* 48:6523–6543
62. Morphy R, Rankovic Z (2006) The physicochemical challenges of designing multiple ligands. *J Med Chem* 49:4961–4970
63. Pe’er D, Hachohen N (2011) Principles and strategies for developing network models in cancer. *Cell* 144:864–873
64. Aittokallio T, Schwikowski B (2006) Graph-based methods for analysing networks in cell biology. *Brief Bioinform* 7:243–255
65. Vidal M, Cusick ME, Barabási AL (2011) Interactome networks and human disease. *Cell* 144:986–998
66. Schadt EE (2009) Molecular networks as sensors and drivers of common human diseases. *Nature* 461:218–223
67. del Sol A, Balling R, Hood L, Galas D (2010) Diseases as network perturbations. *Curr Opin Biotechnol* 21:566–571
68. Chuang H-Y, Hofree M, Ideker T (2012) A Decade of Systems Biology. *Annu Rev Cell Dev Biol* 26:721–744
69. Liu Z-P, Wang Y, Zhang X-S, Chen L (2012) Network-based analysis of complex diseases. *IET Syst Biol* 6:22
70. Barabási A-L, Gulbahce N, Loscalzo J (2011) An Integrative Systems Medicine Approach to Mapping Human Metabolic Diseases. *Nat Rev Genet* 12:56–68
71. Zanzoni A, Soler-López M, Aloy P (2009) A network medicine approach to human disease. *FEBS Lett* 583:1759–1765

72. Creixell P, Schoof EM, Erler JT, Linding R (2012) Navigating cancer network attractors for tumor-specific therapy. *Nat Biotechnol* 30:842–848
73. Leung EL, Cao ZW, Jiang ZH, et al (2013) Network-based drug discovery by integrating systems biology and computational technologies. *Brief Bioinform* 14:491–505
74. Pujol A, Mosca R, Farrés J, Aloy P (2010) Unveiling the role of network and systems biology in drug discovery. *Trends Pharmacol Sci* 31:115–123
75. Arrell DK, Terzic A (2010) Network systems biology for drug discovery. *Clin Pharmacol Ther* 88:120–125
76. Wist AD, Berger SI, Iyengar R (2009) Systems pharmacology and genome medicine: A future perspective. *Genome Med* 1:1–9
77. Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95–108
78. Berger SI, Iyengar R (2009) Network analyses in systems pharmacology. *Bioinformatics* 25:2466–2472
79. Shan Z, Ravi I (2013) Systems Pharmacology: Network Analysis to Identify Multiscale Mechanisms of Drug Action. *Annu Rev Pharmacol Toxicol* 52:505–521
80. Hauser AS, Chavali S, Masuho I, et al (2018) Pharmacogenomics of GPCR Drug Targets. *Cell* 172:41–54
81. Hauser AS, Attwood MM, Rask-andersen M, et al (2017) Trends in GPCR drug discovery : new agents , targets and indications. *Nat Publ Gr* 16:829–842
82. Bautista-Aguilera ÓM, Hagenow S, Palomino-Antolin A, et al (2017) Multitarget-Directed Ligands Combining Cholinesterase and Monoamine Oxidase Inhibition with Histamine H3R Antagonism for Neurodegenerative Diseases. *Angew Chemie - Int Ed* 56:12765–12769
83. Roth S, Kholodenko BN, Smit MJ, Bruggeman FJ (2015) G Protein–Coupled Receptor Signaling Networks from a Systems Perspective. *Mol Pharmacol* 88:604–616
84. Carlsson J, Coleman RG, Setola V, et al (2012) Ligand discovery from a dopamine D-3 receptor homology model and crystal structure. *Nat Chem Biol* 7:769–778
85. Warne T, Serrano-Vega MJ, Baker JG, et al (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454:486–491
86. Munk C, Isberg V, Mordalski S, et al (2016) GPCRdb: the G protein-coupled receptor database – an introduction. *Br J Pharmacol* 173:2195–2207
87. Wacker D, Stevens RC, Roth BL (2017) How Ligands Illuminate GPCR Molecular Pharmacology. *Cell* 170:414–427
88. Kasai RS, Suzuki KGN, Prossnitz ER, et al (2011) Full characterization of GPCR monomer-dimer dynamic equilibrium by single molecule imaging. *J Cell Biol* 192:463–480
89. Rodríguez-Ruiz M, Moreno E, Moreno-Delgado D, et al (2017) Heteroreceptor Complexes Formed by Dopamine D1, Histamine H3, and N-Methyl-D-Aspartate Glutamate Receptors as Targets to Prevent Neuronal Death in Alzheimer’s Disease. *Mol*

90. Marshall FH (2016) Visualizing GPCR ‘Megaplexes’ Which Enable Sustained Intracellular Signaling. *Trends Biochem Sci* 41:985–986
91. Copeland RA (2013) Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists: Second Edition
92. McDonald AG, Tipton KF (2012) Enzymes: Irreversible Inhibition. *eLS* 1–17
93. Ramsay RR, Tipton KF (2017) Assessment of Enzyme Inhibition: A Review with Examples from the Development of Monoamine Oxidase and Cholinesterase Inhibitory Drugs. *Molecules* 22:E1192
94. Dolles D, Decker M (2017) Dual-Acting Compounds Acting as Receptor Ligands and Enzyme Inhibitors. In: *Design of Hybrid Molecules for Drug Development*. pp 1–317
95. Jaiteh M, Zeifman A, Saarinen M, et al (2018) Docking screens for dual inhibitors of disparate drug targets for Parkinson’s disease. *J Med Chem* 12:5269–5278
96. Hellström M, Harvey AR (2014) Cyclic AMP and the regeneration of retinal ganglion cell axons. *Int J Biochem Cell Biol* 56:66–73
97. Raker VK, Becker C, Steinbrink K (2016) The cAMP Pathway as Therapeutic Target in Autoimmune and inflammatory Diseases. *Front Immunol* 7:1–11
98. Rickles RJ, Pierce LT, Giordano TP, et al (2010) Adenosine A_{2A} receptor agonists and PDE inhibitors: a synergistic multitarget mechanism discovered through systematic combination screening in B-cell malignancies. *Blood* 116:593–603
99. Hanaoka K, Guggino WB (2000) cAMP regulates cell proliferation and cyst formation in autosomal polycystic kidney disease cells. *J Am Soc Nephrol* 11:1179–1187
100. Kim MO, Lee YJ, Park JH, et al (2012) PKA and cAMP stimulate proliferation of mouse embryonic stem cells by elevating GLUT1 expression mediated by the NF- κ B and CREB/CBP signaling pathways. *Biochim Biophys Acta - Gen Subj* 1820:1636–1646
101. Vitali E, Peverelli E, Giardino E, et al (2014) Cyclic adenosine 3’-5’-monophosphate (cAMP) exerts proliferative and anti-proliferative effects in pituitary cells of different types by activating both cAMP-dependent protein kinase A (PKA) and exchange proteins directly activated by cAMP (Epac). *Mol Cell Endocrinol* 383:193–202
102. Indolfi C, Avvedimento EV, Di Lorenzo E, et al (1997) Activation of cAMP-PKA signaling in vivo inhibits smooth muscle cell proliferation induced by vascular injury. *Nat Med* 3:775–779
103. Toll L, Jimenez L, Waleh N, et al (2011) β_2 -adrenergic receptor agonists inhibit the proliferation of 1321N1 astrocytoma cells. *J Pharmacol Exp Ther* 336:524–532
104. Rodriguez G, Ross JA, Nagy ZS, Kirken RA (2013) Forskolin-inducible cAMP pathway negatively regulates T-cell proliferation by uncoupling the interleukin-2 receptor complex. *J Biol Chem* 288:7137–7146
105. Kang T-W, Choi SW, Yang S-R, et al (2015) Growth arrest and forced differentiation of human primary glioblastoma multiforme by a novel small molecule. *Sci Rep* 4:5546
106. Sawa A, Chiba T, Ishii J, et al (2017) Effects of sorafenib and an adenylyl cyclase

- activator on in vitro growth of well-differentiated thyroid cancer cells. 64:1115–1123
107. Baillie GS (2009) Compartmentalized signalling: Spatial regulation of cAMP by the action of compartmentalized phosphodiesterases. *FEBS J* 276:1790–1799
 108. Shen S, Wiemelt AP, McMorris FA, Barres BA (1999) Retinal Ganglion Cells Lose Trophic Responsiveness after Axotomy. *Neuron* 23:285–295
 109. Cui Q, Yip HK, Zhao RCH, et al (2003) Intraocular elevation of cyclic AMP potentiates ciliary neurotrophic factor-induced regeneration of adult rat retinal ganglion cell axons. *Mol Cell Neurosci* 22:49–61
 110. Park K, Luo J, Hisheh S, et al (2004) Cellular Mechanisms Associated with Spontaneous and Ciliary Neurotrophic Factor–cAMP-Induced Survival and Axonal Regeneration of Adult Retinal Ganglion Cells. *J Neurosci* 24:10806–10815
 111. Pearse DD, Pereira FC, Marcillo AE, et al (2004) cAMP and Schwann cells promote axonal growth and functional recovery after spinal cord injury. *Nat Med* 10:610–616
 112. Nikulina E, Tidwell JL, Dai HN, et al (2004) The phosphodiesterase inhibitor rolipram delivered after a spinal cord lesion promotes axonal regeneration and functional recovery. *PNAS* 101:8786–8790
 113. Chierzi S, Ratto GM, Verma P, Fawcett JW (2005) The ability of axons to regenerate their growth cones depends on axonal type and age , and is regulated by calcium , cAMP and ERK. *Eur J Neurosci* 21:2051–2062
 114. Costa LM, Pereira JE, Filipe VM, et al (2013) Rolipram promotes functional recovery after contusive thoracic spinal cord injury in rats. *Behav Brain Res* 243:66–73
 115. Gomes C V, Kaster MP, Tomé AR, et al (2011) Adenosine receptors and brain diseases : Neuroprotection and neurodegeneration. *Biochim Biophys Acta* 1808:1380–1399
 116. Stiles GL (1992) Adenosine receptors. *JBiolChem* 267:6451–6454
 117. Schmidt CJ, Chapin DS, Cianfroga J, et al (2008) Preclinical characterization of selective phosphodiesterase 10A inhibitors: a new therapeutic approach to the treatment of schizophrenia. *J Pharmacol Exp Ther* 325:681–690
 118. Giampà C, Patassini S, Borreca A, et al (2009) Phosphodiesterase 10 inhibition reduces striatal excitotoxicity in the quinolinic acid model of Huntington’s disease. *Neurobiol Dis* 34:450–456
 119. Niccolini F, Foltynie T, Reis Marques T, et al (2015) Loss of phosphodiesterase 10A expression is associated with progression and severity in Parkinson’s disease. *Brain* 138:3003–3015
 120. García AM, Brea J, González-García A, et al (2017) Targeting PDE10A GAF Domain with Small Molecules: A Way for Allosteric Modulation with Anti-Inflammatory Effects. *Molecules* 1472:1–8
 121. Oliveira RF, Terrin A, di Benedetto G, et al (2010) The role of type 4 phosphodiesterases in generating microdomains of cAMP: Large scale stochastic simulations. *PLoS One* 5:e11725
 122. Krishnamurthy S, Moorthy BS, Xiang LX, et al (2014) Active site coupling in PDE:PKA complexes promotes resetting of mammalian cAMP signaling. *Biophys J* 107:1426–

123. Marko D, Romanakis K, Zankl H, et al (1998) Induction of Apoptosis by an Inhibitor of cAMP-Specific PDE in Malignant Murine Carcinoma Cells Overexpressing PDE Activity in Comparison to Their Nonmalignant Counterparts. *Cell Biochem Biophys* 28:75–101
124. Miller CL, Cai Y, Oikawa M, et al (2011) Cyclic nucleotide phosphodiesterase 1A: A key regulator of cardiac fibroblast activation and extracellular matrix remodeling in the heart. *Basic Res Cardiol* 106:1023–1039
125. Vandenwijngaert S, Pokreisz P, Hermans H, et al (2013) Increased Cardiac Myocyte PDE5 Levels in Human and Murine Pressure Overload Hypertrophy Contribute to Adverse LV Remodeling. *PLoS One* 8:e58841
126. Ham J, Evans BAJ (2012) An emerging role for adenosine and its receptors in bone homeostasis. *Front Endocrinol (Lausanne)* 3:1–8
127. Rich TC, Fagan KA, Nakata H, et al (2000) Cyclic nucleotide-gated channels colocalize with adenylyl cyclase in regions of restricted cAMP diffusion. *J Gen Physiol* 116:147–161
128. Beltejar M-CG, Lau H-T, Golkowski MG, et al (2017) Analyses of PDE-regulated phosphoproteomes reveal unique and specific cAMP-signaling modules in T cells. *Proc Natl Acad Sci* 114:E6240–E6249
129. Lee K, Lindsey AS, Li N, et al (2015) β -catenin nuclear translocation in colorectal cancer cells is suppressed by PDE10A inhibition, cGMP elevation, and activation of PKG. *Oncotarget* 7(5): 5353–5365.
130. Mediavilla-Varela M, Luddy K, Noyes D, et al (2013) Antagonism of adenosine A_{2A} receptor expressed by lung adenocarcinoma tumor cells and cancer associated fibroblasts inhibits their growth. *Cancer Biol Ther* 14:860–868
131. Inoue Y, Yoshimura K, Kurabe N, et al (2017) Prognostic impact of CD73 and A_{2A} adenosine receptor expression in non-small-cell lung cancer. *Oncotarget* 8:8738–8751
132. Zhu B, Lindsey A, Li N, et al (2017) Phosphodiesterase 10A is overexpressed in lung tumor cells and inhibitors selectively suppress growth by blocking beta-catenin and MAPK signaling. *Oncotarget* 8:69264–69280
133. Mervin LH, Afzal AM, Drakakis G, et al (2015) Target prediction utilising negative bioactivity data covering large chemical space. *J Cheminform* 7:1–16
134. Halgren TA, Murphy RB, Friesner RA, et al (2004) Glide : A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem* 47:1750–1759
135. Zhang W, Pei J, Lai L (2017) Computational Multitarget Drug Design. *J Chem Inf Model* 57:403–412
136. Bottegoni G, Veronesi M, Bisignano P, et al (2016) Development and Application of a Virtual Screening Protocol for the Identification of Multitarget Fragments. *ChemMedChem* 11:1259–1263
137. Méndez-Lucio O, Naveja JJ, Vite-Caritino H, et al (2016) Review. One Drug for

Multiple Targets: A Computational Perspective. *J Mex Chem Soc* 60:168–181

138. Perez-Castillo Y, Sánchez-Rodríguez A, Tejera E, et al (2018) A desirability-based multi objective approach for the virtual screening discovery of broad-spectrum anti-gastric cancer agents. *PLoS One* 13:1–20
139. Wong YH, Lin CL, Chen TS, et al (2015) Multiple target drug cocktail design for attacking the core network markers of four cancers using ligand-based and structure-based virtual screening methods. *BMC Med Genomics* 8:1–23
140. Ma XH, Shi Z, Tan C, et al (2010) In-silico approaches to multi-target drug discovery computer aided multi-target drug design, multi-target virtual screening. *Pharm Res* 27:739–749
141. Nikolic K, Mavridis L, Djikic T, et al (2016) Drug design for CNS diseases: Polypharmacological Profiling of Compounds using Cheminformatic, 3D-QSAR and Virtual Screening Methodologies. *Front Neurosci* 10:1–21
142. Lim H, Gray P, Xie L, Poleksic A (2016) Improved genome-scale multi-target virtual screening via a novel collaborative filtering approach to cold-start problem. *Sci Rep* 6:1–11
143. Thai NQ, Nguyen HL, Linh HQ, Li MS (2017) Protocol for fast screening of multi-target drug candidates: Application to Alzheimer’s disease. *J Mol Graph Model* 77:121–129
144. Katsila T, Spyroulias GA, Patrinos GP, Matsoukas MT (2016) Computational approaches in target identification and drug discovery. *Comput Struct Biotechnol J* 14:177–184
145. Prati F, Cavalli A, Bolognesi ML (2016) Navigating the Chemical Space of Multitarget-Directed Ligands: From Hybrids to Fragments in Alzheimer’s Disease. *Molecules* 21:1–12
146. Costantino L, Barlocco D (2013) Challenges in the design of multitarget drugs against multifactorial pathologies: a new life for medicinal chemistry? *Future Med Chem* 5:5–7
147. Leelananda SP, Lindert S (2016) Computational methods in drug discovery. *Beilstein J Org Chem* 12:2694–2718
148. Sliwoski G, Kothiwale S, Meiler J, Lowe EW (2014) Computational Methods in Drug Discovery. *Pharmacol Rev* 66:334–395
149. Yao T, Xie J, Liu X, et al (2017) Integration of pharmacophore mapping and molecular docking in sequential virtual screening: towards the discovery of novel JAK2 inhibitors. *RSC Adv* 7:10353–10360
150. Marshall GR, Barray CD, Bosshard HE, et al (1979) The Conformational Parameter in Drug Design: The Active Analog Approach. *Comput Drug Des* 112:205–226
151. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
152. Knegtel RMA, Kuntz ID, Oshiro CM (1997) Molecular Docking to Ensembles of Protein Structures. *J Mol Biol* 266:424–440

153. Cho ARTE, Guallar V, Berne BJ, Friesner R (2005) Importance of Accurate Charges in Molecular Docking: Quantum Mechanical/Molecular Mechanical (QM/MM) Approach. *J Comput Chem* 26:915–931
154. Kumar A, Voet A, Zhang KYJ (2012) Fragment Based Drug Design: From Experimental to Computational Approaches. *Curr Med Chem* 19:5128–5147
155. Baurin N, Baker R, Richardson C, et al (2004) Drug-like Annotation and Duplicate Analysis of a 23-Supplier Chemical Database Totalling 2.7 Million Compounds. *J Chem Inf Model* 44:643–651
156. Bautista-Aguilera OM, Samadi A, Chioua M, et al (2014) N-methyl-N-((1-methyl-5-(3-(1-(2-methylbenzyl)piperidin-4-yl)propoxy)-1H-indol-2-yl)methyl)prop-2-yn-1-amine, a new cholinesterase and monoamine oxidase dual inhibitor. *J Med Chem* 57:10455–10463
157. Lepailleur A, Freret T, Boulouard M, et al (2014) Dual Histamine H₃R/Serotonin 5-HT₄R Ligands with Antiamnesic Properties: Pharmacophore-Based Virtual Screening and Polypharmacology. *J Chem Inf Model* 54:1773–1784
158. Ceres N, Ros S, Coccone SS, et al (2009) Discovery of a New Class of Potential Multifunctional Atypical Antipsychotic Agents Targeting Dopamine D₃ and Serotonin 5-HT_{1A} and 5-HT_{2A} Receptors: Design, Synthesis, and Effects on Behavior. *J Med Chem* 52:151–169
159. Lewell XQ, Judd DB, Watson SP, Hann MM (1998) RECAP-Retrosynthetic Combinatorial Analysis Procedure: A Powerful New Technique for Identifying Privileged Molecular Fragments with Useful Applications in Combinatorial Chemistry. *J Chem Inf Comput Sci* 38:511–522
160. Vucicevic J, Srdic-Rajic T, Pieroni M, et al (2016) A combined ligand- and structure-based approach for the identification of rilmenidine-derived compounds which synergize the antitumor effects of doxorubicin. *Bioorganic Med Chem* 24:3174–3183
161. Glen RC, Bender A, Arnby CH, et al (2006) Circular fingerprints: flexible molecular descriptors with applications from physical chemistry to ADME. *IDrugs* 9:199–204
162. Bender A, Glen RC (2004) Molecular similarity: A key technique in molecular informatics. *Org Biomol Chem* 2:3204–3218
163. Willett P, Barnard JM, Downs GM (1998) Chemical similarity searching. *J Chem Inf Comput Sci* 38:983–996
164. Gfeller D, Grosdidier A, Wirth M, et al (2014) SwissTargetPrediction: A web server for target prediction of bioactive small molecules. *Nucleic Acids Res* 42:32–38
165. Mervin LH, Bulusu KC, Kalash L, et al (2018) Orthologue chemical space and its influence on target prediction. *Bioinformatics* 34:72–79
166. Awale M, Reymond J (2017) The polypharmacology browser: a web-based multi-fingerprint target prediction tool using ChEMBL bioactivity data. *J Cheminform* 9:1–10
167. Murphy KP (2006) Naive Bayes classifiers Generative classifiers. University of British Columbia
168. Nidhi, Glick M, Davies JW, Jenkins JL (2006) Prediction of biological targets for

- compounds using multiple-category bayesian models trained on chemogenomics databases. *J Chem Inf Model* 46:1124–1133
169. Olah M, Mracec M, Ostopovici L, et al (2004) WOMBAT: World of Molecular Bioactivity. In: *Chemoinformatics in Drug Discovery*. pp 223–229
 170. Koutsoukas A, Lowe R, Kalantarmotamedi Y, et al (2013) In silico target predictions: Defining a benchmarking data set and comparison of performance of the multiclass Naïve Bayes and Parzen-Rosenblatt Window. *J Chem Inf Model* 53:1957–1966
 171. Gaulton A, Bellis LJ, Bento AP, et al (2011) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40:1–8
 172. Wilson G, Lill M (2011) Integrating structure-based and ligand- based approaches for computational drug design. *Future Med Chem* 3:735–750
 173. Verdonk ML, Cole JC, Hartshorn MJ, et al (2003) Improved protein-ligand docking using GOLD. *Proteins* 52:609–623
 174. Clark KP, Ajay (1995) Flexible ligand docking without parameter adjustment across four ligand–receptor complexes. *J Comput Chem* 16:1210–1226
 175. Taylor JS, Burnett RM (2000) DARWIN: A program for docking flexible molecules. *Proteins Struct Funct Genet* 41:173–191
 176. Wu G, Robertson DH, Brooks CL, Vieth M (2003) Detailed analysis of grid-based molecular docking: A case study of CDOCKER - A CHARMM-based MD docking algorithm. *J Comput Chem* 24:1549–1562
 177. Morris GM, Goodsell DS, Halliday RS, et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662
 178. Goodford PJ (1985) A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J Med Chem* 28:849–857
 179. Kastenholtz MA, Pastor M, Cruciani G, et al (2000) GRID/CPCA: A New Computational Tool To Design Selective Ligands. *J Med Chem* 43:3033–3044
 180. Levitt DG, Banaszak LJ (1992) POCKET: A computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *J Mol Graph* 10:229–234
 181. Laskowski RA (1995) SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 13:323–330
 182. Glaser F, Morris RJ, Najmanovich RJ, et al (2006) A method for localizing ligand binding pockets in protein structures. *Proteins Struct Funct Genet* 62:479–488
 183. Brady GP, Stouten PFW (2000) Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 14:383–401
 184. Mezei M (2003) A new method for mapping macromolecular topography. *J Mol Graph Model* 21:463–472
 185. Hassan NM, Alhossary AA, Mu Y, Kwoh CK (2017) Protein-Ligand Blind Docking

- Using QuickVina-W with Inter-Process Spatio-Temporal Integration. *Sci Rep* 7:1–13
186. Brint AT, Willett P (1987) Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. *J Chem Inf Comput Sci* 27:152–158
 187. Fischer D, Norel R, Wolfson H, Nussinov R (1993) Surface motifs by a computer vision technique: Searches, detection, and implications for protein-ligand recognition. *Proteins Struct Funct Bioinforma* 16:278–292
 188. Norel R, Fischer D, Wolfson HJ, Nussinov R (1994) Molecular surface recognition by a computer vision-based technique. *Protein Eng Des Sel* 7:39–46
 189. Moitessier N, Englebienne P, Lee D, et al (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br J Pharmacol* 153:7–26
 190. Kuntz ID, Blaney JM, Oatley SJ, et al (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 161:269–288
 191. Miller MD, Kearsley SK, Underwood DJ, Sheridan RP (1994) FLOG: A system to select “quasi-flexible” ligands complementary to a receptor of known three-dimensional structure. *J Comput Aided Mol Des* 8:153–174
 192. Diller DJ, Merz KM (2001) High throughput docking for library design and library prioritization. *Proteins Struct Funct Genet* 43:113–124
 193. Burkhard P, Taylor P, Walkinshaw MD (1998) An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 Å X-ray structure of a thrombin-ligand complex. *J Mol Biol* 277:449–466
 194. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 261:470–489
 195. DesJarlais RL, Kuntz ID, Sheridan RP, et al (1986) Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape. *J Med Chem* 29:2149–2153
 196. Leach AR, Kuntz ID (1992) Conformational analysis of flexible ligands in macromolecular receptor sites. *J Comput Chem* 13:730–748
 197. Ewing TJA, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 15:411–428
 198. Welch W, Ruppert J, Jain AN (1996) Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol* 3:449–462
 199. Schnecke V, Kuhn LA (2000) Virtual screening with solvation and ligand-induced complementarity. *Perspect Drug Discov Des* 20:171–190
 200. Zsoldos Z, Reid D, Simon A, et al (2006) eHiTS: An Innovative Approach to the Docking and Scoring Function Problems. *Curr Protein Pept Sci* 7:421–435
 201. Miranker A, Karplus M (1991) Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins Struct Funct Bioinforma* 11:29–34
 202. Eisen MB, Wiley DC, Karplus M, Hubbard RE (1994) HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a

- macromolecule binding site. *Proteins Struct Funct Bioinforma* 19:199–221
203. Böhm HJ (1992) LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 6:593–606
 204. Goodsell DS, Lauble H, Stout CD, Olson AJ (1993) Automated docking in crystallography: Analysis of the substrates of aconitase. *Proteins Struct Funct Bioinforma* 17:1–10
 205. Hart TN, Read RJ (1992) A multiple-start Monte Carlo docking method. *Proteins Struct Funct Bioinforma* 13:206–222
 206. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. *Proteins Struct Funct Bioinforma* 8:195–202
 207. Abagyan R, Totrov M, Kuznetsov D (1994) ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 15:488–506
 208. McMartin C, Bohacek RS (1997) QXP: Powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des* 11:333–344
 209. Luty B, Wasserman Z, Stouten PFW, Hodge CN (1995) A Molecular Mechanics/Grid Method for Evaluation of Ligand-Receptor Interactions. *J Comput Chem* 16:454–464
 210. Jones G, Willett P, Glen RC, et al (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748
 211. Oshiro CM, Kuntz ID, Dixon JS (1995) Flexible ligand docking using a genetic algorithm. *J Comput Aided Mol Des* 9:113–130
 212. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3:935–949
 213. Kollman P (1993) Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem Rev* 93:2395–2417
 214. Aqvist J, Luzhkov VB, Brandsdal BO (2002) Ligand binding affinities from MD simulations. *Acc Chem Res* 35:358–365
 215. Carlson HA, Jorgensen WL (1995) An Extended Linear-Response Method for Determining Free-Energies of Hydration. *J Phys Chem* 99:10667–10673
 216. Shoichet BK, Stroud RM, Santi DV, et al (1993) Structure-based discovery of inhibitors of thymidylate synthase. *Science* 259:1445–1450
 217. Michel J, Verdonk ML, Essex JW (2006) Protein-ligand binding affinity predictions by implicit solvent simulations: A tool for lead optimization? *J Med Chem* 49:7427–7439
 218. Briggs JM, Marrone TJ, McCammon JA (1996) Computational science: New horizons and relevance to pharmaceutical design. *Trends Cardiovasc Med* 6:198–206
 219. Böhm HJ (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J Comput Aided Mol Des* 12:309–323
 220. Gehlhaar DK, Verkhivker GM, Rejto PA, et al (1995) Molecular recognition of the

- inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 2:317–324
221. Verkhivker GM, Bouzida D, Gehlhaar DK, et al (2000) Deciphering common failures in molecular docking of ligand-protein complexes. *J Comput Aided Mol Des* 14:731–751
 222. Jain AN (1996) Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J Comput Aided Mol Des* 10:427–440
 223. Head RD, Smythe ML, Oprea TI, et al (1996) VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J Am Chem Soc* 118:3959–3969
 224. Gehlhaar DK, Moerder KE, Zichi D, et al (1995) De Novo Design of Enzyme Inhibitors by Monte Carlo Ligand Generation. *J Med Chem* 38:466–472
 225. Eldridge MD, Murray CW, Auton TR, et al (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 11:425–445
 226. Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J Med Chem* 42:791–804
 227. Mitchell JBO, Laskowski RA, Alex A, Thornton JM (1999) BLEEP - Potential of mean force describing protein-ligand interactions: I. Generating potential. *J Comput Chem* 20:1165–1176
 228. Ishchenko A V., Shakhnovich EI (2002) SMOG2001 (SMoG2001): An improved knowledge-based scoring function for protein-ligand interactions. *J Med Chem* 45:2770–2780
 229. Feher M, Deretey E, Roy S (2003) BHB: A simple knowledge-based scoring function to improve the efficiency of database screening. *J Chem Inf Comput Sci* 43:1316–1327
 230. Verkhivker G, Appelt K, Freer ST, Villafranca JE (1995) Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng Des Sel* 8:677–691
 231. Wallqvist A, Jernigan RL, Covell DG (1995) A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci* 4:1881–1903
 232. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* 295:337–356
 233. DeWitte RS, Shakhnovich EI (1996) SMOG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J Am Chem Soc* 118:11733–11744
 234. Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 42:5100–5109

235. Feher M (2006) Consensus scoring for protein-ligand interactions. *Drug Discov Today* 11:421–428
236. Clark RD, Strizhev A, Leonard JM, et al (2002) Consensus scoring for ligand/protein interactions. *J Mol Graph Model* 20:281–295
237. Srinivasan J, Cheatham TE, Cieplak P, et al (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J Am Chem Soc* 120:9401–9409
238. Kollman PA, Massova I, Reyes C, et al (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc Chem Res* 33:889–897
239. Clark Still W, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J Am Chem Soc* 112:6127–6129
240. Guimarães CRW, Mathiowetz AM (2010) Addressing limitations with the MM-GB/SA scoring procedure using the watermap method and free energy perturbation calculations. *J Chem Inf Model* 50:547–559
241. Singh N, Warshel A (2010) Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins Struct Funct Bioinforma* 78:1705–1723
242. Ballester PJ, Mitchell JBO (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics* 26:1169–1175
243. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P (2017) Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. *Bioinformatics* 34:3666–3674
244. Gabb HA, Jackson RM, Sternberg MJE (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 272:106–120
245. Bron C, Kerbosch J (1973) Algorithm 457: Finding All Cliques of an Undirected Graph. *Commun ACM* 272:106–120
246. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461
247. Böhm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 8:243–256
248. Teague SJ (2003) Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2:527–541
249. Gschwend DA, Kuntz ID (1996) Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. *J Comput Aided Mol Des* 10:123–132
250. Leach AR (1994) Ligand Docking to Proteins with Discrete Side-chain Flexibility. *J*

251. Desmet J, M DM, Hazes B, Lasters I (1992) The dead-end elimination theorem and its use in protein sidechain positioning. *Nature* 356:539–542
252. Morris GM, Huey R, Lindstrom W, et al (2010) NIH Public Access. *J Comput Chem* 30:2785–2791
253. Carlson HA, Masukawa KM, Rubins K, et al (2000) Developing a Dynamic Pharmacophore Model for HIV-1 Integrase. *J Med Chem* 43:2100–2114
254. Cavasotto CN, Abagyan RA, Llc M, Torrey N (2004) Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J Mol Biol* 337:209–225
255. Zeng L, Zhou M (2002) Bromodomain: an acetyl-lysine binding domain. *FEBS* 513:124–128
256. Venkitakrishnan RP, Zaborowski E, Mcelheny D, et al (2004) Conformational Changes in the Active Site Loops of Dihydrofolate Reductase during the Catalytic Cycle. *Biochemistry* 43:16046–16055
257. Go N, Scheraga HA (1969) Ring Closure and Local Conformational Deformations of Chain Molecules. *Macromolecules* 4537:178–187
258. Dodd LR, Boone TD, Theodorou DN (2006) A concerted rotation algorithm for atomistic Monte Carlo simulation of polymer melts and glasses. *Mol Phys* 78:961–996
259. Parthasarathy S, Murthy MRN (2000) Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng* 13:9–13
260. Cornell WD, Cieplak P, Bayly CI, et al (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 117:5179–5197
261. Weiner SJ, Kollman PA, Case DA, et al (1984) A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J Am Chem Soc* 106:765–784
262. Brooks BR, Bruccoleri RE, Olafson BD, et al (1983) CHARMM - Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* 4:187–217
263. Durrant J, McCammon JA (2011) Molecular dynamics simulations and drug discovery. *BMC Biol* 9:1–9
264. Wang J, Wolf RM, Caldwell JW, et al (2004) Development and Testing of a General Amber Force Field. *J Comput Chem* 25:1157–1174
265. Bowers K, Chow E, Xu H, et al (2006) Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In: *Proceedings of the ACM/IEEE SC 2006 Conference (SC'06)*. pp 11–17
266. Adcock SA, Mccammon JA (2006) Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem Rev* 106:1589–1615
267. Verlet L (1967) Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* 159:98–103
268. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian

- equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 23:327–341
269. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A Linear Constraint Solver for Molecular Simulations BERK. *J Comput Chem* 18:1463–1472
 270. Lambrakos SG, Boris JP, Oran ES, et al (1989) A modified shake algorithm for maintaining rigid bonds in molecular dynamics simulations of large molecules. *J Comput Phys* 85:473–486
 271. Anézo C, de Vries AH, Höltje H-D, et al (2003) Methodological Issues in Lipid Bilayer Simulations. *J Phys Chem B* 107:9424–9433
 272. Allen MP, Tildesley DJ (2017) *Computer simulation of liquids: Second edition*
 273. Berendsen HJC, Postma JPM, Van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3690
 274. Uline MJ, Corti DS (2013) Molecular dynamics at constant pressure: Allowing the system to control volume fluctuations via a “shell” particle. *Entropy* 15:3941–3969
 275. White JA, Román FL, González A, Velasco S (2008) Periodic boundary conditions and the correct molecular-dynamics ensemble. *Phys A Stat Mech its Appl* 387:6705–6711
 276. van Gunsteren WF, Berendsen HJC (1990) *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry*. *Angew Chemie Int Ed English* 29:992–1023
 277. Bakowies D, Baron R, Christen M, et al (2005) The GROMOS Software for Biomolecular Simulation : *J Comput Chem* 26:1719–1751
 278. Harder E, Damm W, Maple J, et al (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J Chem Theory Comput* 12:281–296
 279. Wildman J, Repiščák P, Paterson MJ, Galbraith I (2016) General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution. *J Chem Theory Comput* 12:3813–3824
 280. Tironi IG, Sperb R, Smith PE, Van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. *J Chem Phys* 102:5451–5459
 281. Essmann U, Perera L, Berkowitz ML, et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593
 282. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
 283. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W : improving the sensitivity of progressive multiple sequence alignment through sequence weighting , position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
 284. Kandt C, Schlitter J, Gerwert K (2004) Dynamics of Water Molecules in the Bacteriorhodopsin Trimer in Explicit Lipid/Water Environment. *Biophys J* 86:705–717
 285. Róg T, Murzyn K, Gurbel R, et al (2004) Effects of phospholipid unsaturation on the bilayer nonpolar region. *J Lipid Res* 45:326–336

286. Allen TW, Baştuğ T, Kuyucak S, Chung SH (2003) Gramicidin A channel as a test ground for molecular dynamics force fields. *Biophys J* 84:2159–2168
287. Ng HW, Laughton CA, Doughty SW (2014) Molecular dynamics simulations of the adenosine A_{2a} receptor in POPC and POPE lipid bilayers: Effects of membrane on protein behavior. *J Chem Inf Model* 54:573–581
288. Keränen H, Gutiérrez-de-Terán H, Åqvist J (2014) Structural and energetic effects of A_{2A} adenosine receptor mutations on agonist and antagonist binding. *PLoS One* 9(10):e108492
289. Kubinyi H (2006) *Computer Applications in Pharmaceutical Research and Development*
290. Kroemer RT (2007) *Structure-Based Drug Design: Docking and Scoring*. *Curr Protein Pept Sci* 8:312–328
291. Venhorst J, Ter Laak AM, Commandeur JNM, et al (2003) Homology modeling of rat and human cytochrome P450 2D (CYP2D) isoforms and computational rationalization of experimental ligand-binding specificities. *J Med Chem* 46:74–86
292. Williams P a, Cosme J, Ward A, et al (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* 424:464–468
293. Meng X-Y, Zheng Q-C, Zhang H-X (2009) A comparative analysis of binding sites between mouse CYP2C38 and CYP2C39 based on homology modeling, molecular dynamics simulation and docking studies. *Biochim Biophys Acta* 1794:1066–1072
294. Boehm HJ, Boehringer M, Bur D, et al (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J Med Chem* 43:2664–2674
295. Doman TN, McGovern SL, Witherbee BJ, et al (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* 45:2213–2221
296. Bautista-Aguilera OM, Esteban G, Chioua M, et al (2014) Multipotent cholinesterase/monoamine oxidase inhibitors for the treatment of Alzheimer's disease: Design, synthesis, biochemical evaluation, ADMET, molecular modeling, and QSAR analysis of novel donepezil-pyridyl hybrids. *Drug Des Devel Ther* 8:1893–1910
297. Bolea I, Juárez-Jiménez J, De Los Ríos C, et al (2011) Synthesis, biological evaluation, and molecular modeling of donepezil and N-[(5-(Benzyloxy)-1-methyl-1H-indol-2-yl)methyl]-N-methylprop-2-yn-1-amine hybrids as new multipotent cholinesterase/monoamine oxidase inhibitors for the treatment of Alzheimer's disease. *J Med Chem* 54:8251–8270
298. Cherkasov A, Muratov EN, Fourches D, et al (2015) QSAR modeling: where have you been? Where are you going to? *J Med Chem* 57:4977–5010
299. Danishuddin, Khan AU (2016) Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discov Today* 21:1291–1302
300. Ismaili L, Refouvelet B, Benchekroun M, et al (2017) Multitarget compounds bearing tacrine- and donepezil-like structural and functional motifs for the potential treatment of Alzheimer's disease. *Prog Neurobiol* 151:4–34

301. Cramer RD (2012) The inevitable QSAR renaissance. *J Comput Aided Mol Des* 26:35–38
302. Aniceto N, Freitas AA, Bender A, Ghafourian T (2016) A novel applicability domain technique for mapping predictive reliability across the chemical space of a QSAR: Reliability-density neighbourhood. *J Cheminform* 8:1–20
303. Pajouhesh H, Lenz GR (2005) Medicinal chemical properties of successful central nervous system drugs. *NeuroRx* 2:541–553
304. Jeffrey P, Summer S (2010) Neurobiology of Disease Assessment of the blood – brain barrier in CNS drug discovery. *Neurobiol Dis J* 37:33–37
305. Rankovic Z (2015) CNS Drug Design: Balancing Physicochemical Properties for Optimal Brain Exposure. *J Med Chem* 58:2584–2608
306. Bingham M, Rankovic Z (2012) Chapter 18: medicinal chemistry challenges in CNS drug discovery. *Drug discovery for psychiatric disorders*. The Royal Society of Chemistry, Cambridge, pp 465–509
307. Hitchcock SA (2012) Structural modifications that alter the P-glycoprotein efflux properties of compounds. *J Med Chem* 55:4877–4895
308. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 6:881–890
309. Leeson PD, Davis AM (2004) Time-related differences in the physical property profiles of oral drugs. *J Med Chem* 47:6338–6348
310. Weiss MM, Williamson T, Babu-Khan S, et al (2012) Design and preparation of a potent series of hydroxyethylamine containing β -secretase inhibitors that demonstrate robust reduction of central β -amyloid. *J Med Chem* 55:9009–9024
311. Alelyunas YW, Empfield JR, McCarthy D, et al (2010) Experimental solubility profiling of marketed CNS drugs, exploring solubility limit of CNS discovery candidate. *Bioorganic Med Chem Lett* 20:7312–7316
312. Schrödinger Release 2018-2: QikProp
313. Lagorce D, Sperandio O, Baell JB, et al (2015) FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res* 43:W200–W207
314. Ghose AK, Herbertz T, Hudkins RL, et al (2012) Knowledge-based, central nervous system (CNS) lead selection and lead optimization for CNS drug discovery. *ACS Chem Neurosci* 3:50–68
315. Desai P V., Sawada GA, Watson IA, Raub TJ (2013) Integration of in silico and in vitro tools for scaffold optimization during drug discovery: Predicting P-glycoprotein efflux. *Mol Pharm* 10:1249–1261
316. Siklos M, BenAissa M, Thatcher GRJ (2015) Cysteine proteases as therapeutic targets: Does selectivity matter? A systematic review of calpain and cathepsin inhibitors. *Acta Pharm Sin B* 5:506–519
317. Ono Y, Sorimachi H (2012) Calpains- An elaborate proteolytic system. *Biochim Biophys Acta - Proteins Proteomics* 1824:224–236

318. Low KE, Partha SK, Davies PL, Campbell RL (2014) Allosteric inhibitors of calpains: Reevaluating inhibition by PD150606 and LSEAL. *Biochim Biophys Acta* 1840:3367–3373
319. Lebon G, Warne T, Edwards PC, et al (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. *Nature* 474:521–525
320. Lebon G, Edwards PC, Leslie AGW, Tate CG (2015) Molecular Determinants of CGS21680 Binding to the Human Adenosine A_{2A} Receptors. *Mol Pharmacol* 87:907–915
321. Carpenter B, Lebon G (2017) Human Adenosine A_{2A} Receptor: Molecular Mechanism of Ligand Binding and Activation. *Front Pharmacol* 8:1–15
322. Li Y, Bondada V, Joshi A, Geddes JW (2009) Calpain 1 and Calpastatin Expression is Developmentally Regulated. *Exp Neurol* 220:316–319
323. Huang Y, Wang KKW (2001) The calpain family and human disease. *Trends Mol Med* 7:355–362
324. Ono Y, Saido TC, Sorimachi H (2016) Calpain research for drug discovery: challenges and potential. *Nat Rev Drug Discov* 15:854–876
325. Kurbatskaya K, Phillips EC, Croft CL, et al (2016) Upregulation of calpain activity precedes tau phosphorylation and loss of synaptic proteins in Alzheimer's disease brain. *Acta Neuropathol Commun* 4:1–15
326. Shinohara K, Tomioka M, Nakano H, et al (1996) Apoptosis induction resulting from proteasome inhibition. *Biochem J* 317:385–388
327. Wang KK, Nath R, Posner A, et al (1996) An alpha-mercaptoacrylic acid derivative is a selective nonpeptide cell-permeable calpain inhibitor and is neuroprotective. *Proc Natl Acad Sci U S A* 93:6687–6692
328. Todd B, Moore D, Deivanayagam CCS, et al (2003) A structural model for the inhibition of calpain by calpastatin: Crystal structures of the native domain VI of calpain and its complexes with calpastatin peptide and a small molecule inhibitor. *J Mol Biol* 328:131–146
329. Adams SE, Parr C, Miller DJ, et al (2012) Potent inhibition of Ca²⁺-dependent activation of calpain-1 by novel mercaptoacrylates. *Med Chem Commun* 3:566–570
330. Adams SE, Rizkallah PJ, Miller DJ, et al (2014) The structural basis of differential inhibition of human calpain by indole and phenyl α -mercaptoacrylic acids. *J Struct Biol* 187:236–241
331. Kumar P, Choonara YE, Pillay V (2015) In silico affinity profiling of neuroactive polyphenols for post-traumatic calpain inactivation: A molecular docking and atomistic simulation sensitivity analysis. *Molecules* 20:135–168
332. Lu L, Meehan MJ, Gu S, et al (2015) Mechanism of action of thalassospiramides, a new class of calpain inhibitors. *Sci Rep* 5:1–8
333. <https://www.aldrichmarketselect.com/> (accessed 2016-10-10)
334. Madhavi Sastry G, Adzhigirey M, Day T, et al (2013) Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided*

335. Shelley JC, Cholleti A, Frye LL, et al (2007) Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des* 21:681–691
336. Wang Y, Xiao J, Suzek TO, et al (2012) PubChem's BioAssay Database. *Nucleic Acids Res* 40:D400-12
337. R Core Team. R: A Language and Environment for Statistical Computing (version 3.2.4). (2016).
338. Kalash L, Val C, Azuaje J, et al (2017) Computer-aided design of multi-target ligands at A1R, A2AR and PDE10A, key proteins in neurodegenerative diseases. *J Cheminform* 9:1–19
339. Baell JB, Holloway G a. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53:2719–2740
340. Capuzzi SJ, Muratov EN, Tropsha A (2017) Phantom PAINS : Problems with the Utility of Alerts for Pan-Assay INterference CompoundS. *J Chem Inf Model* 57:417–427
341. ChemAxon Standardizer. <https://www.chemaxon.com/products/standardizer> (accessed 2017-02-27)
342. Berthold M, Cebon N, Dill F (2008) KNIME: The Konstanz information miner. *SIGKDD Explor* 11:26–31
343. Hadley Wickham (2009) ggplot2: elegant graphics for data analysis
344. SigmaPlot (Systat Software, San Jose, CA)
345. Polizzi KM, Kontoravdi C (2013) Combinatorial DNA Assembly Using Golden Gate Cloning. , *Synthetic Biology, Methods in Molecular Biology (Methods and Protocols)*, vol. 1073, Humana Press, Totowa, NJ
346. Busse WW, Wenzel SE, Meltzer EO, et al (2013) Safety and efficacy of the prostaglandin D₂ receptor antagonist AMG 853 in asthmatic patients. *J Allergy Clin Immunol* 131:339–345
347. Silver K, Leloup L, Freeman LC, et al (2010) Non-steroidal Anti-inflammatory Drugs Inhibit Calpain Activity and Membrane Localization of Calpain 2 Protease. *Int J Biochem Cell Biol* 42:2030–2036
348. Tong-Jun L (2009) Inhibition of calpain reduces allergic inflammation. 029751.0 A:1–7
349. Shook BC, Rassnick S, Wallace N, et al (2012) Design and characterization of optimized adenosine A_{2A}/A₁ receptor antagonists for the treatment of Parkinson's disease. *J Med Chem* 55:1402–17
350. Mondal C, Halder AK, Adhikari N, Jha T (2014) Structural findings of cinnolines as anti-schizophrenic PDE10A inhibitors through comparative chemometric modeling. *Mol Divers* 18:655–671
351. Knott EP, Assi M, Rao SNR, et al (2017) Phosphodiesterase Inhibitors as a Therapeutic

- Approach to Neuroprotection and Repair. *Int J Mol Sci* 18:1–38
352. Molecular Operating Environment (MOE), 2013, Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2016
 353. Kovacs GG (2014) Current Concepts of Neurodegenerative Diseases. *Eur Med J* 1:78–86
 354. Lang AE, Obeso JA (2004) Personal view Challenges in Parkinson's disease: restoration of the nigrostriatal dopamine system is not enough. *Lancet Neurol* 3:309–316
 355. Latini S, Pedata F (2001) Adenosine in the central nervous system: Release mechanisms and extracellular concentrations. *J Neurochem* 79:463–484
 356. Li M, Wang X, Meintzer MKAY, et al (2000) Cyclic AMP Promotes Neuronal Survival by Phosphorylation of Glycogen Synthase Kinase β . *Mol Cell Biol* 20:9356–9363
 357. Weber L (2002) The application of multi-component reactions in drug discovery. *Curr Med Chem* 9:2085–2093
 358. Musonda CC, Gut J, Rosenthal PJ, et al (2006) Application of multicomponent reactions to antimalarial drug discovery. Part 2: New antiplasmodial and antitrypanosomal 4-aminoquinoline γ - and δ -lactams via a 'catch and release' protocol. *Bioorganic Med Chem* 14:5605–5615
 359. Jacobson KA, Gao Z-G (2006) Adenosine receptors as therapeutic targets. *Nat Rev Drug Discov* 5:247–264
 360. Piirainen H, Ashok Y, Nanekar RT, Jaakola V (2011) Structural features of adenosine receptors : From crystal to function. *Biochim Biophys Acta* 1808:1233–1244
 361. Kolosionek E, Savai R, Ghofrani HA, et al (2009) Expression and Activity of Phosphodiesterase Isoforms during Epithelial Mesenchymal Transition : The Role of Phosphodiesterase 4. *Mol Biol Cell* 20:4751–4765
 362. Katritch V, Kufareva I, Abagyan R (2012) Structure Based Prediction of Subtype-Selectivity for Adenosine Receptor Antagonists. *Neuropharmacology* 60:108–115 .
 363. Glukhova A, Thal DM, Nguyen AT, et al (2017) Structure of the Adenosine A₁ Receptor Reveals the Basis for Subtype Selectivity. *Cell* 168:867–877
 364. Sander T, Freyss J, von Korff M, Rufener C (2015) DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. *J Chem Inf Model* 55:460–473
 365. RDKit: Cheminformatics and Machine Learning Software, 2013. [http:// www.rdkit.org](http://www.rdkit.org). (accessed 2017-02-27)
 366. Liggi S, Drakakis G, Hendry AE, et al (2013) Extensions to in silico bioactivity predictions using pathway annotations and differential pharmacology analysis: Application to xenopus laevis phenotypic readouts. *Mol Inform* 32:1009–1024
 367. Jones E, Oliphant T, Peterson P SciPy: Open Source Scientific Tools for Python, 2001. <http://www.scipy.org/> (accessed 2017-02-27)
 368. Liu W, Chun E, Thompson AA, et al (2012) Structural Basis for Allosteric Regulation

- of GPCRs by Sodium Ions. *Science* 337:232–236
369. Hu E, Kunz RK, Rumfelt S, et al (2012) Discovery of potent, selective, and metabolically stable 4-(pyridin-3-yl) cinnolines as novel phosphodiesterase 10A (PDE10A) inhibitors. *Bioorg Med Chem Lett* 22:2262–2265
 370. Yaziji V, Rodriguez D, Gutierrez-de-Terran H, et al (2011) Pyrimidine Derivatives as Potent and Selective A₃ Adenosine Receptor Antagonists. *J Med Chem* 54:457–471
 371. Iii CLB, Dixon JS, Moulton J, Stevens RC (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat Rev Drug Discov* 8:455–463
 372. Schrödinger Release 2016-4: LigPrep, Schrödinger, LLC, New York, NY, 2016
 373. de Lera Ruiz M, Lim Y-H, Junying Z (2013) Adenosine A_{2A} Receptor as a Drug Discovery Target. *J Med Chem* 57:3623–3650
 374. Drabczyńska A, Yuzlenko O, Köse M, et al (2011) Synthesis and biological activity of tricyclic cycloalkylimidazo-, pyrimido- and diazepinopurinediones. *Eur J Med Chem* 46:3590–3607
 375. Khunnawutmanotham, Nisachon Sahakitpichan P, Chimnoi N, Techasakul S (2015) Divergent Total Syntheses to Azafluoranthene and Dehydroaporphine Alkaloids. *European J Org Chem* 28:6324–6332
 376. Wagner S, Scheunemann M, Dipper K, Egerland, Ute, Hoefgen, Nobert, Steinbach, Jorg, Burst P (2016) Development of highly potent phosphodiesterase 10A (PDE10A) inhibitors: Synthesis and in vitro evaluation of 1,8-dipyridinyl- and 1-pyridinyl-substituted imidazo [1, 5-a] quinoxalines. *Eur J Med Chem* 107:97–108
 377. Chen P, Doweyko AM, Norris D, et al (2004) Imidazoquinoxaline Src-Family Kinase p56^{Lck} Inhibitors: SAR, QSAR, and the Discovery of (S)-N-(2-Chloro-6-methylphenyl)-2-(3-methyl-1-piperazinyl)imidazo-[1,5-a]pyrido[3,2-e]pyrazin-6-amine (BMS-279700) as a Potent and Orally Active Inhibitor with Excellent in Vivo Antiinflammatory Activity. *J Med Chem* 47:4517–4529
 378. Bartolome JM, Diego SAA De, Artola M, et al (2015) Identification of a Novel Orally Bioavailable Phosphodiesterase 10A (PDE10A) Inhibitor with Efficacy in Animal Models of Schizophrenia. *J Med Chem* 58:978–993
 379. Malamas MS, Ni Y, Erdei J, et al (2011) Highly Potent, Selective, and Orally Active Phosphodiesterase 10A Inhibitors. *J Med Chem* 54:7621–7638
 380. Neustadt BR, Hao J, Lindo N, et al (2007) Potent, selective, and orally active adenosine A_{2A} receptor antagonists: Arylpiperazine derivatives of pyrazolo[4,3-e]-1,2,4-triazolo[1,5-c]pyrimidines. *Bioorg Med Chem Lett* 17:1376–1380
 381. Bolcato C, Cusan C, Pastorin G, et al (2008) Pyrazolo-triazolo-pyrimidines as adenosine receptor antagonists: Effect of the N-5 bond type on the affinity and selectivity at the four adenosine receptor subtypes. *Purinergic Signal* 4:39–46
 382. Kehler J, Ritzen A, Langgård M, et al (2011) Triazoloquinazolines as a novel class of phosphodiesterase 10A (PDE10A) inhibitors. *Bioorg Med Chem Lett* 21:3738–3742
 383. Zadpour M, Behbahani FK (2015) Iron (III) phosphate as a green and reusable catalyst

- for the synthesis of 4,6-disubstituted 2-aminopyridine-3-carbonitriles. *Monatshefte für Chemie - Chem Mon* 146:1865–1869
384. Patoliya PU, Gohel VP, Purohit DM, Patolia VN (2015) Synthesis and biological evaluation of some new cyano pyridine derivatives. *J Chem Pharm Res* 7:182–186
 385. Bosch MP, Campos F, Niubo I, et al (2004) Synthesis and Biological Activity of New Potential Agonists for the Human Adenosine A_{2A} Receptor. *J Med Chem* 47:4041–4053
 386. Shipe WD, Sharik SS, Barrow JC, et al (2015) Discovery and Optimization of a Series of Pyrimidine-Based Phosphodiesterase 10A (PDE10A) Inhibitors through Fragment Screening, Structure-Based Design, and Parallel Synthesis. *J Med Chem* 58:7888–7894
 387. Motulsky H, Christopoulos A (2004) Fitting Models to Biological Data using Linear and Nonlinear Regression. A Practical Guide to Curve Fitting.
 388. Zolfigol M, Kiafar M, Yarie M, Taherpour A (2016) Experimental and theoretical studies of the nanostructured {Fe₃O₄@SiO₂@(CH₂)₃Im}C(CN)₃ catalyst for 2-amino-3-cyanopyridine preparation *via* an anomeric based oxidation. *RSC Adv* 6:50100–50111
 389. Khalili D (2016) Graphene oxide : a reusable and metal-free carbocatalyst for the one-pot synthesis of 2-amino-3-cyanopyridines in water. *Tetrahedron Lett* 57:1721–1723
 390. Mantri M, Graaf O De, Veldhoven J Van, et al (2008) 2-Amino-6-furan-2-yl-4-substituted Nicotinonitriles as A_{2A} Adenosine Receptor Antagonists. *J Med Chem* 51:4449–4455
 391. Safari J, Hossein S, Dehghan S (2012) Ultrasound-promoted an efficient method for one-pot synthesis of 2-amino-4,6-diphenylnicotinonitriles in water: A rapid procedure without catalyst. *Ultrason Sonochem* 19:1061–1069
 392. Lingam VSP, Dahale, Dnyaneshwar H, Rathi, Vijay E, et al (2015) Design, Synthesis, and Pharmacological Evaluation of 5,6-Disubstituted Pyridin-2(1H) -one Derivatives as Phosphodiesterase 10A (PDE10A) Antagonists. *J Med Chem* 58:8292–8308
 393. Abadi AH, Ibrahim TM, Abouzid KM, et al (2009) Design, synthesis and biological evaluation of novel pyridine derivatives as anticancer agents and phosphodiesterase 3 inhibitors. *Bioorg Med Chem* 17:5974–5982
 394. Meegalla SK, Huang H, Illig CR, et al (2016) Discovery of novel potent imidazo [1, 2-b] pyridazine PDE10a inhibitors. *Bioorg Med Chem Lett* 26:4216–4222
 395. Raheem IT, Schreier JD, Fuerst J, et al (2016) Discovery of pyrazolopyrimidine phosphodiesterase 10A inhibitors for the treatment of schizophrenia. *Bioorg Med Chem Lett* 26:126–132
 396. Yoshikawa M, Hitaka T, Hasui T, et al (2016) Design and synthesis of potent and selective pyridazin-4(1H)-one-based PDE10A inhibitors interacting with Tyr683 in the PDE10A selectivity pocket. *Bioorg Med Chem* 24:3447–3455
 397. Kuhn B, Guba W, Banner D, et al (2016) A Real-World Perspective on Molecular Design. *J Med Chem* 59:4087–4102
 398. Underwood JG, Bragg AE, Neumark DM, et al (2008) The 2.6 Angstrom Crystal Structure of a Human A_{2A} Adenosine Receptor Bound to an Antagonist. *Science* 322:1211–1218

399. Segala E, Guo D, Cheng RKY, et al (2016) Controlling the Dissociation of Ligands from the Adenosine A_{2A} Receptor through Modulation of Salt Bridge Strength. *J Med Chem* 59:6470–6479
400. Knight A, Hemmings JL, Win I, et al (2016) Discovery of Novel Adenosine Receptor Agonists That Exhibit Subtype Selectivity. *J Med Chem* 59:947–964
401. Nguyen ATN, Baltos J, Thomas T, et al (2016) Extracellular Loop 2 of the Adenosine A₁ Receptor Has a Key Role in Orthosteric Ligand Affinity and Agonist Efficacy. *Mol Pharmacol* 90:703–714
402. Halder AK, Amin SA, Jha T, Gayen S (2017) Insight into the structural requirements of pyrimidine-based phosphodiesterase 10A (PDE10A) inhibitors by multiple validated 3D QSAR approaches. *SAR QSAR Environ Res* 28:253–273
403. Yuan G, Gedeon NG, Jankins TC, Jones GB (2015) Novel approaches for targeting the adenosine A_{2A} receptor. *Expert Opin Drug Discov* 10:63–80
404. Rodríguez D, Gao ZG, Moss SM, et al (2015) Molecular docking screening using agonist-bound GPCR structures: Probing the A_{2A} adenosine receptor. *J Chem Inf Model* 55:550–563
405. Xu F, Wu H, Katritvh V, et al (2011) Structure of an Agonist-Bound Human A_{2A} Adenosine Receptor. *Science* 332:322–327
406. P RE, FS M, Leite AA, et al (2017) New Antitumoral Pharmacological Strategies Involving Ca²⁺/cAMP Signaling Pathways. *J Cancer Epidemiol Prev* 2:1–6
407. Fajardo AM, Piazza GA, Tinsley HN (2014) The Role of Cyclic Nucleotide Signaling Pathways in Cancer: Targets for Prevention and Treatment. *Cancers* 6:436–458
408. Fusco JP, Pita G, Pajares MJ, et al (2018) Genomic characterization of individuals presenting extreme phenotypes of high and low risk to develop tobacco-induced lung cancer. *Cancer Med* 7: 3474-3483
409. He R, Li X, Liang L, et al (2017) The suppressive role of miR-542-5p in NSCLC: the evidence from clinical data and in vivo validation using a chick chorioallantoic membrane model. *BMC Cancer* 17:1–15
410. Rieger JM, Brown ML, Sullivan GW, et al (2001) Design, Synthesis, and Evaluation of Novel A_{2A} Adenosine Receptor Agonists. *J Med Chem* 44:531–539
411. Chen J, Liu M, Chern T, et al (2011) Design and Synthesis of Novel Dual-Action Compounds Targeting the Adenosine A_{2A} Receptor and Adenosine Transporter for Neuroprotection. *ChemMedChem* 6:1390–1400
412. Pourbasheer E, Tabar SS, Masand VH, et al (2015) 3D-QSAR and docking studies on adenosine A_{2A} receptor antagonists by the CoMFA method. *SAR QSAR Environ Res* 26:461–472
413. Hu E, Kunz RK, Rumfelt S, et al (2012) Use of structure based design to increase selectivity of pyridyl-cinnoline phosphodiesterase 10A (PDE10A) inhibitors against phosphodiesterase 3 (PDE3). *Bioorg Med Chem Lett* 22:6938–6942
414. Deganutti G, Moro S (2017) Supporting the identification of novel fragment-based positive allosteric modulators using a supervised molecular dynamics approach: A

- retrospective analysis considering the human A_{2A} adenosine receptor as a key example. *Molecules* 22:1–12
415. McGraw C, Robinson AS (2017) Membrane Cholesterol and the Adenosine A_{2a} Receptor. *Biophys J* 112:33a–34a
 416. Yuan S, Hu Z, Filipek S, Vogel H (2015) W2466.48 opens a gate for a continuous intrinsic water pathway during activation of the adenosine A_{2A} receptor. *Angew Chemie - Int Ed* 54:556–559
 417. Guo D, Pan AC, Dror RO, et al (2016) Molecular Basis of Ligand Dissociation from the Adenosine A_{2A} Receptor. *Mol Pharmacol* 89:485–491
 418. Caliman AD, Swift SE, Wang Y, et al (2013) Investigation of the Conformational Dynamics of the A_{2A} Adenosine Receptor by Molecular Dynamics Simulation. *Mol Biophys* 58:618–634
 419. Watson ME a, Marzioch M, Clare JJ, et al (2000) Functional coupling of mammalian receptors to the yeast mating pathway using novel yeast /mammalian G protein α -subunit chimeras. *Yeast* 16:11–22
 420. Ng HW, Laughton CA, Doughty SW (2013) Molecular dynamics simulations of the adenosine A_{2a} receptor: Structural stability, sampling, and convergence. *J Chem Inf Model* 53:1168–1178
 421. Irwin JJ, Shoichet BK (2005) ZINC - A free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182
 422. Congreve M, Andrews SP, Dore AS, et al (2012) Discovery of 1,2,4-Triazine Derivatives as Adenosine A_{2A} Antagonists using Structure Based Drug Design. *J Med Chem* 55:1898–1903
 423. Nicola GF De, Martin S, Bullard B, Pastore A (2010) Solution Structure of the Apo C-Terminal Domain of the Lethocerus F1 Troponin C Isoform. *Biochemistry* 49:1719–1726
 424. Weinert T, Olieric N, Cheng R, et al (2017) Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat Commun* 8:1–11
 425. Melnikov I, Polovinkin V, Kovalev K, et al (2017) Fast iodide-SAD phasing for high-throughput membrane protein structure determination. *Sci Adv* 3:1–12
 426. Sun B, Bachhawat P, Chu ML, et al (2017) Crystal structure of the adenosine A_{2A} receptor bound to an antagonist reveals a potential allosteric pocket. *PNAS* 114:2–7
 427. Lomize MA, Pogozheva ID, Joo H, et al (2012) OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res* 40:370–376
 428. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual Molecular Dynamics. *J Mol Graph* 14:33–38
 429. seaborn. In: <https://seaborn.pydata.org>
 430. Ladds G, Zervou S, Vatish M, et al (2009) Regulators of G protein signalling proteins in the human myometrium. *Eur J Pharmacol* 610:23–28
 431. Spandidos A, Wang X, Wang H, Seed B (2010) PrimerBank : a resource of human and

- mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res* 38:792–799
432. Katritch V, Jaakola V, Lane JR, et al (2010) Structure-Based Discovery of Novel Chemotypes for Adenosine A_{2A} Receptor Antagonists. *J Med Chem* 53:1799–1809
 433. Jaakola VP, Lane JR, Lin JY, et al (2010) Ligand binding and subtype selectivity of the human A_{2A} adenosine receptor: Identification and characterization of essential amino acid residues. *J Biol Chem* 285:13032–13044
 434. Weston C, Poyner D, Patel V, et al (2014) Investigating G protein signalling bias at the glucagon-like peptide-1 receptor in yeast. *Br J Pharmacol* 171:3651–3665
 435. Yuan S, Wu R, Latek D, et al (2013) Lipid Receptor S1P(1) Activation Scheme Concluded from Microsecond All-Atom Molecular Dynamics Simulations. *PLoS Comput Biol* 9:1–8
 436. Petrone P, Pande VS (2006) Can conformational change be described by only a few normal modes? *Biophys J* 90:1583–1593
 437. Weston C, Winfield I, Harris M, et al (2016) Receptor activity-modifying protein-directed G protein signaling specificity for the calcitonin gene-related peptide family of receptors. *J Biol Chem* 291:21925–21944
 438. Roberts MJ, Broome RE, Kent TC, et al (2018) The inhibition of human lung fibroblast proliferation and differentiation by Gs-coupled receptors is not predicted by the magnitude of cAMP response. *Respir Res* 19:1–13
 439. Li N, Lee K, Xi Y, et al (2015) Phosphodiesterase 10A: a novel target for selective inhibition of colon tumor cell growth and β -catenin-dependent TCF transcriptional activity. *Oncogene* 34:1499–1509
 440. Perez-Aso M, Fernandez P, Mediero A, et al (2014) Adenosine 2A receptor promotes collagen production by human fibroblasts *via* pathways involving cyclic AMP and AKT but independent of Smad2/3. *FASEB J* 28:802–812
 441. Palmer T, Gettys T, Jacobson K, Stiles G (1994) Desensitization of the canine A_{2a} adenosine receptor: delineation of multiple processes. *Mol Pharmacol* 45:1062–1094
 442. Rosethorne EM, Bradley ME, Kent TC, Charlton SJ (2015) Functional desensitization of the β 2 adrenoceptor is not dependent on agonist efficacy. *Pharmacol Res Perspect* 3:e00101
 443. Mundell S, Kelly E (2011) Adenosine receptor desensitization and trafficking. *Biochim Biophys Acta - Biomembr* 1808:1319–1328
 444. Liu X, Ostrom R, Insel P (2004) cAMP-elevating agents and adenylyl cyclase overexpression promote an antifibrotic phenotype in pulmonary fibroblasts. *AJP Cell Physiol* 286:C1089-1099
 445. Baker JG, Hall IP, Hill SJ (2004) Temporal Characteristics of cAMP Response Element-Mediated Gene Transcription: Requirement for Sustained cAMP Production. *Mol Pharmacol* 65:986–998
 446. Lazarus KA, Hadi F, Zambon E, et al (2018) BCL11A interacts with SOX2 to control the expression of epigenetic regulators in lung squamous cell carcinoma. *Nat Commun*

9: 3327

447. Kalash L, Cresser-Brown J, Habchi J, et al (2018) Structure-based design of allosteric calpain-1 inhibitors populating a novel bioactivity space. *Eur J Med Chem* 157:1264–1275.

Appendix A

The Chemscore is represented in Equation A1, where the first term C_0 is the offset determined from the regression analysis, the second term sums up the ligand-atom/receptor-atom pairs, which are defined as lipophilic. The third term accounts for the ligand-receptor hydrogen-bonding interactions. The f, g, and h functions may result in a full score of 1.00 for distances or angles that lie within nominal limits, and a partial score within 0.00 and 1.00 for distances or angles lying outside these limits but within larger threshold values. For instance, if the H...X hydrogen bond distance is within 0.25 Å of a nominal value of 1.85 Å, then $g(\Delta r)$ would be 1.00 but linearly decreases to zero if the distance is between 2.10 and 2.50 Å. Likewise, $h(\Delta R)$ would be 1.00 if the Z-H...X angle lies within 30° of 180°, and falls to zero if it is between 120° and 150°.

$$\Delta G_{bind} = C_0 + C_{lipo} \sum f(r_{1r}) + C_{hbond} \sum g(\Delta r)h(\Delta \alpha) + C_{metal} \sum f(r_{1m}) + C_{rota}H_{rota} \text{ (Equation A1)}$$

Equation A2 represents GlideScore, where the lipophilic-lipophilic and hydrogen-bonding terms are the same as in ChemScore. However, the hydrogen-bonding term is split into various weighted parts, which are determined by the charges of the donor and acceptor that could be either both neutral (most stable), or at least one or both are charged (least stable). The fifth term, which accounts for metal-ligand interaction, is similar to the term in ChemScore but differs in three major ways. First, it has been modified to consider only interactions with anionic acceptor atoms such as either of the two oxygens of a carboxylate group. In metalloproteases, for example, this would allow for preference to be made in the coordination to metal centers by the anionic ligand functionality. Secondly, in the case where two or more metal ligations are found, Glide would count only the best interaction. Thirdly, the total charge on the metal ion is assessed in the apo structure of the protein, and then if it is positive, the scoring is modified by incorporating preference for an anionic ligand.

$$\begin{aligned} \Delta G_{bind} = & C_{lipo-lipo} \sum f(r_{1r}) + \\ & C_{hbond-neut-neut} \sum g(\Delta r)h(\Delta \alpha) + C_{hbond-neut-charged} \sum g(\Delta r)h(\Delta \alpha) + \\ & + C_{hbond-charged-charged} \sum g(\Delta r)h(\Delta \alpha) + C_{max-metal-ion} \sum f(r_{1m}) + \\ & C_{rota}H_{rota} + C_{polar-phob}V_{polar-phob} + C_{coul}E_{coul} + C_{vdW}E_{vdW} + \\ & \text{solvation terms (Equation A2)} \end{aligned}$$

Appendix B

Equation B1 illustrates the Taylor series that would account for the position of an atom after a short and finite interval of time, denoted by Δt , given the standard position with respect to a single component of vector r_i , (that is the position along a single dimension, x) at a specific time, t .

$$x(t + \Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{d^2x(t)}{dt^2} \frac{\Delta t^2}{2} + \dots \text{ (Equation B1)}$$

In order to account for higher order terms in the Taylor series, an approximation is made where the position $x(t)$, the velocity $dx(t)/dt$, and the acceleration $d^2x(t)/dt^2$ account for a numerical solution of the equations of motion. Equation B2 is Newton's second law for this single dimension, which describes the acceleration, where F_x is the component of the net force acting on the atom parallel to the x direction.[263]

$$\frac{d^2x(t)}{dt^2} = \frac{F_x}{m} \text{ (Equation B2)}$$

Appendix C

Equation C1 illustrates the functional form of the OPLS force field. It includes harmonic terms (the subscript eq denotes the equilibrium bond length r and angle θ) that represent bond stretch and angle bending energies. A truncated Fourier series summing up all dihedral angles represents the torsional terms, and summing Coulomb and Lennard-Jones terms that cover pairwise interactions between interacting sites, denoted by i and j , represent the non-bonded energies. The energy is scaled by f_{ij} for intramolecular atom pairs separated by 3 bonds or less, and for each ij pair, the Lennard-Jones σ and ϵ parameters are derived from the atom site-specific parameters coupled according to the geometric combining rule. [278]

$$E = \sum_{i < j} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right] f_{ij} + \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \left[\frac{V_1}{2} (1 + \cos\varphi) + \frac{V_2}{2} (1 - \cos 2\varphi) + \frac{V_3}{2} (1 + \cos 3\varphi) + \frac{V_4}{2} (1 - \cos 4\varphi) \right]$$

(Equation C1)

The OPLS3 fitting protocol incorporates many parameters to ensure the largest extent of coverage for the medicinal chemistry space of small molecules, where the experimental data is not available in the majority of cases.[278] The protocol was described in detail by Harder *et al.*, [278] The general fitting protocol is highlighted as follows:

- (1) The stretching and bend terms are determined as the first step in the protocol as they are insensitive to the other terms in the force field, and where they are fit to quantum chemical data.
- (2) The van der Waals terms are found from liquid state simulations, along with the core charge set employed in these simulations.
- (3) Defining and testing of bond charge corrections (BCC's) for the CM1A-BCC charge model (a model based on combining Cramer-Truhlar CM1A charges with specifically fit BCC's), is done using, for instance, calculations of solvation free energies in water.
- (4) Finally, the fitting of torsional parameters to quantum chemical data is done in the last step.

This protocol could be used to model compounds, where no experimental data is reported provided that the van der Waals parameters are well accounted for by the force field. The predictive power of the protocol, however, takes into consideration a set of assumptions which has been discussed extensively by Harder *et al.* [278]

Appendix D

```
mysql -u<yourusername> -p pidgin -e "select cid_information.stdsmls from cid_uniprot left  
join cid_information on cid_uniprot.cid = cid_information.cid where cid_uniprot.uniprot =  
'<youruniprot>';" > ~/output.txt
```


Appendix E

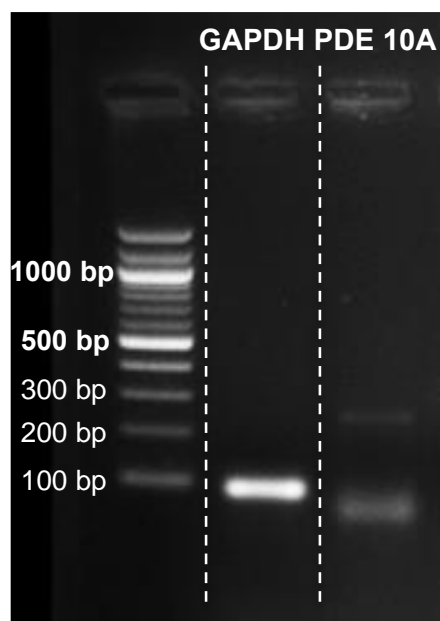


Figure E1. Products obtained (by IW) from RT-PCR using gene specific primers for GAPDH and PDE10A on cDNA. These were produced from RNA extracted from CHO-K1 cells, then shown in representative ($n = 3$) agarose gel. The RT-PCR analysis reveals that CHO-K1 cells endogenously express PDE10A.